

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 16:44:40 ; Search time 9.59302 Seconds  
(without alignments)  
46.007 Million cell updates/sec

Title: US-10-010-667a-2\_COPY\_14\_28  
Perfect score: 86  
Sequence: 1 WKKKPRNLLEDDYL 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/ECTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	86	100.0	339	4 US-09-323-873A-2	Sequence 2, Appl
2	42.5	49.4	256	4 US-09-134-001C-5278	Sequence 5278, Ap
3	40	46.5	221	4 US-09-247-373B-54	Sequence 54, Appl
4	38	44.2	345	1 US-08-403-866-2	Sequence 2, Appl
5	38	44.2	968	3 US-08-651-999A-7	Sequence 7, Appl
6	38	44.2	968	4 US-09-385-752-7	Sequence 7, Appl
7	38	44.2	3969	4 US-08-061-376-5	Sequence 5, Appl
8	37	43.0	25	1 US-08-468-709B-16	Sequence 16, Appl
9	37	43.0	25	2 US-08-241-664B-16	Sequence 16, Appl
10	37	43.0	323	2 US-08-969-106-4	Sequence 4, Appl
11	37	43.0	416	1 US-08-252-995D-2	Sequence 2, Appl
12	37	43.0	416	2 US-08-324-108-2	Sequence 2, Appl
13	37	43.0	464	1 US-08-252-995D-6	Sequence 6, Appl
14	37	43.0	464	2 US-08-834-108-6	Sequence 6, Appl
15	37	43.0	577	1 US-08-484-105-24	Sequence 24, Appl
16	37	43.0	577	2 US-08-484-106-24	Sequence 24, Appl
17	37	43.0	577	1 US-08-756-317-13	Sequence 13, Appl
18	37	43.0	605	4 US-09-394-645-2	Sequence 2, Appl
19	37	43.0	605	4 US-09-243-560B-2	Sequence 2, Appl
20	37	43.0	711	4 US-08-961-083-82	Sequence 82, Appl
21	37	43.0	751	4 US-09-402-929-2	Sequence 2, Appl
22	37	43.0	752	4 US-09-402-929-6	Sequence 6, Appl
23	37	43.0	925	1 US-08-252-995D-4	Sequence 4, Appl
24	37	43.0	925	2 US-08-834-108-4	Sequence 4, Appl
25	37	43.0	1228	4 US-09-605-785-537	Sequence 537, App
26	37	43.0	1228	4 US-09-439-313-537	Sequence 537, App
27	36	41.9	202	2 US-08-286-819A-6	Sequence 6, Appl

28	36	41.9	202	3 US-08-980-357-6	Sequence 6, Appl
29	36	41.9	373	4 US-09-359-161-3	Sequence 3, Appl
30	36	41.9	437	4 US-09-134-001C-3418	Sequence 3418, Ap
31	36	41.9	440	3 US-09-052-778-13	Sequence 13, Appl
32	36	41.9	449	4 US-09-134-001C-5673	Sequence 5673, Ap
33	36	41.9	550	1 US-08-484-493-2	Sequence 2, Appl
34	36	41.9	550	2 US-08-484-494-2	Sequence 2, Appl
35	36	41.9	550	1 US-08-345-212-2	Sequence 2, Appl
36	36	41.9	550	4 US-09-249-003-2	Sequence 4, Appl
37	36	41.9	746	4 US-09-149-934-4	Sequence 4, Appl
38	36	41.9	1805	1 US-07-853-913-2	Sequence 2, Appl
39	36	41.9	1854	4 US-09-004-838-108	Sequence 108, App
40	36	41.9	2291	2 US-08-286-819A-29	Sequence 29, Appl
41	36	41.9	2291	4 US-08-980-357-29	Sequence 29, Appl
42	35.5	41.3	273	4 US-09-651-656-7	Sequence 7, Appl
43	35.5	41.3	273	4 US-09-650-855-7	Sequence 7, Appl
44	35.5	41.3	414	5 PCT-US92-06840-2	Sequence 2, Appl
45	35	40.7	22	1 US-08-484-635-148	Sequence 148, App

## ALIGNMENTS

```
RESULT 1
US-09-323-873A-2
; Sequence 2, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hudert
; APPLICANT: Kahan Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; FILE REFERENCE: 129.160SU2
; CURRENT APPLICATION NUMBER: US/09/323,873A
; PRIOR FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-01
; PRIOR APPLICATION NUMBER: 60/091,183
; PRIOR FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-323-873A-2
Query Match 100.0%; Score 86; DB 4; Length 339;
Best Local Similarity 100.0%; Pred. No. 9.5e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 1 WKKKPRNLLEDDYL 15
Db 14 WKKKPRNLLEDDYL 28
RESULT 2
US-09-134-001C-5278
; Sequence 5278, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
```

PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 5278  
LENGTH: 256  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5278

Query Match 49.4%; Score 42.5; DB 4; Length 256;  
Best Local Similarity 64.3%; Pred. No. 12;  
Matches 9; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

QY 2 KMKPRRNLEEDYL 15  
DB 234 KVKPRNN-KEDNYL 246

RESULT 3  
US-09-247-373B-54  
Sequence 54, Application US/09247373B  
Patent No. 6168954  
GENERAL INFORMATION:  
APPLICANT: MCGONIGLE, BRIAN  
APPLICANT: O'KEEFE, DANIEL  
TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES  
FILE REFERENCE: CL-1108-A  
CURRENT APPLICATION NUMBER: US/09/247,373B  
CURRENT FILING DATE: 1999-02-10  
PRIOR APPLICATION NUMBER: 08/924,747  
PRIOR FILING DATE: 1997-09-05  
NUMBER OF SEQ ID NOS: 56  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 54  
LENGTH: 221  
TYPE: PRT  
ORGANISM: SOYBEAN  
US-09-247-373B-54

Query Match 46.5%; Score 40; DB 4; Length 221;  
Best Local Similarity 44.4%; Pred. No. 27;  
Matches 8; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

QY 1 WKMK---PRRNLEEDYL 14  
DB 18 WTLKXKDIPYONIEDRY 35

RESULT 4  
US-08-403-866-2  
Sequence 2, Application US/08403866  
Patent No. 5643779  
GENERAL INFORMATION:  
APPLICANT: Enrich, Stanislaw  
APPLICANT: Godon, Jean-Jacques  
APPLICANT: Renault, Pierre  
TITLE OF INVENTION: Nucleic acid coding for an alpha-acetolactate  
TITLE OF INVENTION: synthase from Lactococcus and its applications  
NUMBER OF SEQUENCES: 16  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/403,866  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldman, Michael L.  
REGISTRATION NUMBER: 30,727  
REFERENCE/DOCKET NUMBER: 20747/30  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (716) 263-1600  
TELEFAX: (716) 263-1487  
TELEX: 978450 (WUT)  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 345 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Lactococcus lactis subsp. lactis  
INDIVIDUAL ISOLATE: LEUB  
US-08-403-866-2

Query Match 44.2%; Score 38; DB 1; Length 345;  
Best Local Similarity 50.0%; Pred. No. 94;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 KMKPRRNLEEDYL 15  
DB 114 KLSPIRVNVDPL 127

RESULT 5  
US-08-651-999A-7  
Sequence 7, Application US/08651999A  
Patent No. 6031088  
GENERAL INFORMATION:  
APPLICANT: Stefan Somlo and Toshio Mochizuki  
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE PKD2 GENE  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSER: AMSTER, ROTHSTEIN & EBENSTEIN  
STREET: 90 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE  
MEDIUM TYPE: DISKETTE  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/651,999A  
FILING DATE: MAY 23, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: ELIZABETH A. BOGOSIAN  
REGISTRATION NUMBER: 39,911  
REFERENCE/DOCKET NUMBER: 96700/395  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-5995  
TELEFAX: (212) 286-0854 or 286-0082  
TELEX: TWX 710-581-4766  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 968 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE:  
DESCRIPTION: peptide  
HYPOTHETICAL: NO  
US-08-651-999A-7

Query Match 44.2%; Score 38; DB 3; Length 968;  
Best Local Similarity 46.2%; Pred. No. 2,9e+02;  
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 WKMKPRRNLEEDD 13

Db 293 WKQPSNOTEDN 305

RESULT 6  
US-09-385-752-7

Sequence 7, Application US/09385752

Patent No. 6228591

GENERAL INFORMATION:

APPLICANT: Stefan Somlo and Toshio Mochizuki

TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE PKD2 GENE

TITLE OF INVENTION: AND USES THEREOF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN

STREET: 90 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: U.S.A.

ZIP: 10016

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE

MEDIUM TYPE: DISKETTE

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09385,752

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/651,999

FILING DATE: MAY 23, 1996

ATTORNEY/AGENT INFORMATION:

NAME: ELIZABETH A. BOSOSIAN

REGISTRATION NUMBER: 39,911

REFERENCE/DOCKET NUMBER: 96700/395

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 697-5595

TELEFAX: (212) 286-0854 or 286-0082

TELEX: TWX 710-581-4766

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 968 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE:

DESCRIPTION: peptide

HYPOTHETICAL: NO

US-09-385-752-7

Query Match

Best Local Similarity 44.2%; Score 38; DB 4; Length 968;

Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 WKMKPRNLEDD 13

Db 293 WKQPSNOTEDN 305

RESULT 7

US-08-061-376-5

Sequence 5, Application US/08061376

Patent No. 6175000

GENERAL INFORMATION:

APPLICANT: Evans, Glen A.

APPLICANT: Djibali, Malek

APPLICANT: Seltzer, Lucia

APPLICANT: Parry, Pauline

TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11Q23

TITLE OF INVENTION: TRANSLATION BREAKPOINT ASSOCIATED WITH ACUTE LEUKEMIAS

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark

STREET: 444 South Flower Street, Suite 2000

CITY: Los Angeles

STATE: California

COUNTRY: USA

ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/468,709B

FILING DATE: 06/06/95

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Smith, Deann F.

REGISTRATION NUMBER: 36,683

REFERENCE/DOCKET NUMBER: 2115-00676DVF

TELECOMMUNICATION INFORMATION:

TELEPHONE: (810) 641-1600

ADDRESS: Pretty, Schroeder, Brueggemann & Clark

STREET: 444 South Flower Street, Suite 2000

CITY: Los Angeles

STATE: California

COUNTRY: USA

ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/061,376

FILING DATE: 13-MAY-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Reiter, Stephen E.

REGISTRATION NUMBER: 31,192

REFERENCE/DOCKET NUMBER: P41 9387

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619)546-4737

TELEFAX: (619)546-9392

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 3969 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-061-376-5

Query Match

Best Local Similarity 44.2%; Score 38; DB 4; Length 3969;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 5 PRRLLEDDY 14

Db 424 PRRLLEDDY 433

RESULT 8

US-08-468-709B-16

Sequence 16, Application US/08468709B

Patent No. 5654137

GENERAL INFORMATION:

APPLICANT: Astrom, Anders

APPLICANT: Voornes, John

APPLICANT: Patterson, Ulrika

APPLICANT: Tavares, Amir

TITLE OF INVENTION: HUMAN CRABP-I AND CRABP-II

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: Harness, Dickey & Pierce, P.L.C.

STREET: PO Box 828

CITY: Bloomfield Hills

STATE: Michigan

COUNTRY: United States of America

ZIP: 48303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/468,709B

FILING DATE: 06/06/95

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Smith, Deann F.

REGISTRATION NUMBER: 36,683

REFERENCE/DOCKET NUMBER: 2115-00676DVF

TELECOMMUNICATION INFORMATION:

TELEPHONE: (810) 641-1600

TELEFAX: (810) 641-0270  
TELEX: 287637  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-468-709B-16

Query Match 43.0%; Score 37; DB 1; Length 25;  
Best Local Similarity 54.5%; Pred. No. 7.9;  
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 WKMKRRNLEE 11  
DB 7 WKRRSSSENFEE 17

RESULT 9

US-08-241-664B-16  
Sequence 16, Application US/08241664B  
Patent No. 5871909  
GENERAL INFORMATION:  
APPLICANT: Voorhees, John J.  
APPLICANT: Astrom, Anders  
APPLICANT: Patterson, Ulrika  
APPLICANT: Tavaakol, Amir  
TITLE OF INVENTION: HUMAN CRABP-I AND CRABP-II  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.  
STREET: PO Box 828  
CITY: Bloomfield Hills  
STATE: Michigan  
COUNTRY: United States of America  
ZIP: 48303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/241,664B  
FILING DATE: May 11, 1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, Deann F.  
REGISTRATION NUMBER: 36,683  
REFERENCE/DOCKET NUMBER: 2115-00676COD  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (810) 641-1600  
TELEFAX: (810) 641-0270  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-241-664B-16

Query Match 43.0%; Score 37; DB 2; Length 25;  
Best Local Similarity 54.5%; Pred. No. 7.9;  
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 WKMKRRNLEE 11  
DB 7 WKRRSSSENFEE 17

RESULT 10  
US-08-969-106-4

Sequence 4, Application US/08969106  
Patent No. 5986055  
GENERAL INFORMATION:  
APPLICANT: Yang, M.  
APPLICANT: Mandapalan, K.  
APPLICANT: Schultz, V.  
TITLE OF INVENTION: CDK2 INTERACTIONS  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/969,106  
FILING DATE: 13-NOV-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mirock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7934-057  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 323 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-969-106-4

Query Match 43.0%; Score 37; DB 2; Length 323;  
Best Local Similarity 50.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 KPRNLEEDDYL 15  
DB 292 KRRGYEDDYV 303

RESULT 11  
US-08-252-995D-2  
Sequence 2, Application US/08252995D  
Patent No. 5650501  
GENERAL INFORMATION:  
APPLICANT: Dennis, James W.  
APPLICANT: Heffernan, Mike  
APPLICANT: Fode, Carol  
TITLE OF INVENTION: NOVEL SERINE/THREONINE KINASE  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BERSKIN & PARR  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:



APPLICATION NUMBER: US/08/252,995D  
FILING DATE: 02-JUN-1994  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Kurd/dyk, Linda M  
REGISTRATION NUMBER: 34,971  
REFERENCE/DOCKET NUMBER: 3153-96  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 416 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-252-995D-2

Query Match 43.0%; Score 37; DB 1; Length 416;  
Best Local Similarity 54.5%; Pred. No. 1.7e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 4 KPRRLLEDDY 14  
DB 406 KPRRLSDENQH 416

RESULT 12

US-08-834-108-2  
Sequence 2, Application US/08834108  
Patent No. 5976893

GENERAL INFORMATION:  
APPLICANT: Dennis, James W  
APPLICANT: Heffernan, Mike  
APPLICANT: Fode, Carol  
TITLE OF INVENTION: NOVEL SERINE/THREONINE KINASE  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BERSKIN & PARR  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/834,108  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Kurd/dyk, Linda M  
REGISTRATION NUMBER: 34,971  
REFERENCE/DOCKET NUMBER: 3153-210  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 416 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-834-108-2

Query Match 43.0%; Score 37; DB 2; Length 416;  
Best Local Similarity 54.5%; Pred. No. 1.7e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
OY 4 KPRRLLEDDY 14

DB 406 KPRRLSDENQH 416

RESULT 13  
US-08-252-995D-6  
Sequence 6, Application US/08252995D  
Patent No. 5650501

GENERAL INFORMATION:  
APPLICANT: Dennis, James W  
APPLICANT: Heffernan, Mike  
APPLICANT: Fode, Carol  
TITLE OF INVENTION: NOVEL SERINE/THREONINE KINASE  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BERSKIN & PARR  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/252,995D  
FILING DATE: 02-JUN-1994  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Kurd/dyk, Linda M  
REGISTRATION NUMBER: 34,971  
REFERENCE/DOCKET NUMBER: 3153-96  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 464 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-252-995D-6

Query Match 43.0%; Score 37; DB 1; Length 464;  
Best Local Similarity 54.5%; Pred. No. 1.9e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 4 KPRRLLEDDY 14  
DB 406 KPRRLSDENQH 416

RESULT 14

US-08-834-108-6  
Sequence 6, Application US/08834108  
Patent No. 5976893

GENERAL INFORMATION:  
APPLICANT: Dennis, James W  
APPLICANT: Heffernan, Mike  
APPLICANT: Fode, Carol  
TITLE OF INVENTION: NOVEL SERINE/THREONINE KINASE  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BERSKIN & PARR  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/834,108  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Kurdvdyk, Linda M  
REGISTRATION NUMBER: 34,971  
REFERENCE/DOCKET NUMBER: 3153-210  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 464 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-834-108-6

Query Match 43.0%; Score 37; DB 2; Length 464;  
Best Local Similarity 54.3%; Pred. No. 1.9e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 4 KPRRNLEEDY 14  
DB 406 KPRSLDENQH 416

## RESULT 15

US-08-484-105-24  
Sequence 24, Application US/08484105

Patent No. 5589341

GENERAL INFORMATION:

APPLICANT: STILLMAN, Bruce

APPLICANT: BELL, Stephen P

APPLICANT: KOBAYASHI, Ryuji

APPLICANT: RINE, Jasper

APPLICANT: FOSS, Margit

APPLICANT: MCNALLY, Francis J

APPLICANT: LAURENSEN, Patricia

APPLICANT: HERSKOWITZ, Ira

APPLICANT: LI, Joachim J

APPLICANT: GAVIN, Kimberly

TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESSES:

ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/484,105

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Osman Ph. D., Richard Aron

REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 494-8700

TELEFAX: (415) 494-8771

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 24:

SEQUENCE CHARACTERISTICS:  
LENGTH: 577 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-484-105-24

Query Match 43.0%; Score 37; DB 1; Length 577;  
Best Local Similarity 70.0%; Pred. No. 2.4e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 4 KPRRNLEEDD 13  
DB 52 KPEYDLEEDD 61

Search completed: March 26, 2003, 16:51:37  
Job time: 11.593 secs

GenCore version 5.1.4 p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: March 26, 2003, 16:38:14 ; Search time 29.1279 Seconds  
(without alignments)  
68.620 Million cell updates/sec

Title: US-10-010-667A-2\_COPY\_14\_28  
Perfect score: 86  
Sequence: 1 WKMKPRRLERDDYL 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A.Geneseq\_101002:\*

1:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1980.DAT:*
2:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1981.DAT:*
3:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1982.DAT:*
4:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1983.DAT:*
5:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1984.DAT:*
6:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1985.DAT:*
7:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1986.DAT:*
8:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1987.DAT:*
9:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1988.DAT:*
10:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1989.DAT:*
11:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1990.DAT:*
12:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1991.DAT:*
13:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1992.DAT:*
14:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1993.DAT:*
15:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1994.DAT:*
16:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1995.DAT:*
17:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1996.DAT:*
18:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1997.DAT:*
19:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1998.DAT:*
20:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA1999.DAT:*
21:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA2000.DAT:*
22:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA2001.DAT:*
23:	/SID2/gcgdata/geneeq/geneeqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	86	100.0	15	21	AAV58201 Human STRAP-1 pept
2	86	100.0	15	22	AAE02784 Immunogenic peptid
3	86	100.0	66	20	AAV12305 Human 5' EST seque
4	86	100.0	109	20	AAV12304 Human 5' EST seque
5	86	100.0	339	21	AAV58194 Human STRAP-1 prot
6	86	100.0	339	22	AAU69927 Human prostate CDN
7	86	100.0	339	22	AAW78845 Human protein SER
8	86	100.0	339	22	AAW01282 P789P amino acid s
9	86	100.0	339	23	ABG61813 Prostate cancer-as
10	86	100.0	339	23	ABB95387 Human P789P protei

11	86	100.0	374	22	AAW79829 Human protein SBQ
12	86	100.0	375	22	AAE02780 Human six transmem
13	70	81.4	95	20	AAV11840 Human 5' EST seque
14	63	73.3	339	20	AAW66309 Kidney injury asso
15	45	52.3	72	22	ABG25517 Novel human diagno
16	45	52.3	421	22	ABG25522 Novel human diagno
17	45	52.3	1106	22	ABG25533 Novel human diagno
18	43	50.0	2040	23	ABB04597 Lettuce big-vein v
19	42.5	49.4	256	23	ABP40433 Novel human diagno
20	42	48.8	156	22	ABG04393 Staphylococcus epi
21	42	48.8	164	21	AAW35757 EIAV rev gene prod
22	42	48.8	164	22	AAW67883 Equine infectious
23	42	48.8	165	22	AAW67877 Equine infectious
24	42	48.8	472	22	AAU03543 Human protein Xina
25	42	48.8	959	22	AAU07874 Polypeptide sequen
26	42	48.8	996	23	AAU75900 Human Testis speci
27	42	48.8	1019	22	AAU07892 Polypeptide sequen
28	42	48.8	1417	22	ABW63329 Drosophila melanog
29	41.5	48.3	208	22	AAU18520 Human cytoskeletal
30	41	47.7	166	21	AAW30007 Arabidopsis thalia
31	41	47.7	192	21	AAW30006 Arabidopsis thalia
32	41	47.7	271	21	AAW30005 Arabidopsis thalia
33	41	47.7	327	22	AAW40194 Human polypeptide
34	41	47.7	327	23	ABW06781 Human cancer cell
35	41	47.7	738	22	ABW65092 Drosophila melanog
36	40	46.5	161	22	AAU01103 Gene 41 Human secr
37	40	46.5	197	20	AAW98006 Human interleukin-
38	40	46.5	205	21	AAW38584 Arabidopsis thalia
39	40	46.5	221	21	AAW07850 Amino acid sequenc
40	40	46.5	223	22	AAW62225 Glycine max glutat
41	40	46.5	224	21	AAW34829 Arabidopsis thalia
42	40	46.5	224	21	AAW39745 Arabidopsis thalia
43	40	46.5	236	23	ABP28691 Streptococcus poly
44	40	46.5	257	23	ABP30273 Streptococcus poly
45	40	46.5	258	23	ABP25968 Streptococcus poly

## ALIGNMENTS

RESULT 1  
AAV58201  
ID AAV58201 standard; peptide; 15 AA.  
XX  
AC AAV58201;  
XX  
DT 14-MAR-2000 (first entry)  
XX  
DE Human STRAP-1 peptide 1, corresponding to STRAP-1 residues 14-28.  
XX  
KW Serpentine transmembrane antigen of the prostate; STRAP-1; prostate;  
transmembrane domain; type IIIa membrane protein; expression; cancer;  
prostate cancer; bladder cancer; colon cancer; pancreatic cancer;  
ovarian cancer; tumour antigen; immunisation; immune response;  
cellular; humoral; anticancer vaccine; antibody; diagnosis;  
prognosis; monitoring; susceptibility; therapeutic inhibitor;  
drug targeting; recombinant protein.  
XX  
OS Synthetic.  
OS Homo sapiens.  
XX  
PN WO9962941-A2.  
XX  
PD 09-DEC-1999.  
XX  
PF 01-JUN-1999; 99WO-US12157.  
XX  
PR 01-JUN-1998; 98US-0087520.  
PR 30-JUN-1998; 98US-0091183.  
XX  
PA (UROC-) UROGENESYS INC.  
PA (AFAR/) AFAR D E.  
PA (HUBB/) HUBERT R S.

PA (LEON/) LEONG K.  
 PA (RAIT/) RAITANO A. B.  
 PA (SAFR/) SAFRAN D. C.  
 XX  
 PI Afar DE, Hubert RS, Leong K, Raitano AB, Safiran DC,  
 XX MPI, 2000-072832/06.  
 XX  
 PT Novel proteins useful as diagnostic markers and therapeutic targets,  
 particularly for prostatic cancer -  
 XX  
 PS Example 3C; Page 41; 83pp; English.  
 XX  
 CC This sequence represents a synthetic peptide (peptide 1) corresponding  
 CC to residues 14-28 of STRAP-1 (serpentine transmembrane antigen of  
 CC the prostate, AAY58199). This peptide was used to raise  
 CC polyclonal anti-STRAP-1 N-terminus antibodies in sheep. The  
 CC antibodies were used to probe Western blots of cell lysates from  
 CC prostate and non-prostate cancer cell lines, and in immunohistochemical  
 CC analysis of STRAP-1 expression. STRAP-1 is the prototype  
 CC member of the STRAP family of proteins (AAY58194-Y58197) which  
 CC exhibit a high degree of structural conservation, but which show  
 CC no significant structural homology to known human proteins. The STRAP-1  
 CC gene has been localised to chromosome 7p22. STRAP-1 is thought to be a  
 CC type IIIa membrane protein and is expressed predominantly in prostate  
 CC cells in normal human tissues. Structurally, STRAP-1 is a 339 amino  
 CC acid protein characterised by six transmembrane domains and  
 CC intracellular N- and C-termini, suggesting that it folds in a  
 CC "serpentine" manner into three extracellular and two intracellular loops.  
 CC STRAP-1 mRNA and protein expression is maintained at high levels and  
 CC throughout all stages of prostate cancer. STRAP-1 mRNA and/or protein is  
 CC also overexpressed in certain other cancers, including bladder, colon,  
 CC pancreatic and ovarian cancer. The function of the STRAP proteins is not  
 CC known. They may be ion channels (from the presence of six transmembrane  
 CC domains), a feature which is shared by certain ion channels) or  
 CC gap-junction proteins (from immunohistochemical staining). STRAP-1 and  
 CC STRAP-2 are cell-surface tumour antigens. Immunisation with a STRAP  
 CC protein induces cellular and humoral immune responses against  
 CC STRAP-expressing cells. STRAP proteins may be used to identify  
 CC specific-binding agents, to produce anticancer vaccines and to generate  
 CC specific antibodies. The antibodies may be used for detection, prognosis,  
 CC and monitoring of cancers (or susceptibility to cancer), as therapeutic  
 CC inhibitors or to target therapeutic agents to their site of action. STRAP  
 CC nucleic acids may be used for recombinant protein production, as  
 CC diagnostic and prognostic reagents, for identifying STRAP-expressing  
 CC cells for screening inhibitors of STRAP expression and for therapeutic  
 CC modulation/inhibition of STRAP expression. Since high levels of STRAP  
 CC proteins are exposed on the cell surface, they are easily targeted by  
 CC systemically administered agents, and because they are expressed mainly  
 CC on prostatic epithelial cells, agents targeted to them should have  
 CC minimal side effects on other tissues.  
 CC  
 XX  
 SQ Sequence 15 AA;  
 Query Match 100.0%; Score 86; DB 21; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKMKPRRLLEDDYL 15  
 DB 1 WKMKPRRLLEDDYL 15

RESULT 2  
 ID AAE02784 standard; peptide; 15 AA.  
 XX  
 AC AAE02784;  
 XX  
 DT 06-AUG-2001 (first entry)  
 XX  
 DE Immunogenic peptide #1 of human STEAP-1.  
 XX

KW Human; cytostatic; antiproliferative; vaccine; gene therapy;  
 KW six transmembrane epithelial antigen of the prostate-1; STEAP-1;  
 KW chromosome 7p22.3; cancer; prostate; colon; bladder; pancreatic;  
 KW ovarian; lung; serpentine transmembrane antigen; immunogen.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200140276-A2.  
 XX  
 XX 07-JUN-2001.  
 XX  
 XX 06-DEC-2000; 2000MO-US33040.  
 XX  
 PR 06-DEC-1999; 99US-0455486.  
 XX  
 PA (UROG-) UROGENESIS INC.  
 XX  
 PI Afar DEH, Hubert RS, Raitano AB, Safiran DC, Mitchell SC, Faris M,  
 XX Jakobovits A;  
 XX MPI, 2001-367804/38.  
 XX  
 PT New STEAP (six transmembrane epithelial antigen of the prostate)  
 PT proteins, expressed in human cancers, useful for detecting and treating  
 PT cancer -  
 XX  
 PS Example 3C; Page 77; 187pp; English.  
 XX  
 CC The present invention relates to human six transmembrane epithelial  
 CC antigen of the prostate (STEAP) protein. STEAP is a member of cell  
 CC surface serpentine transmembrane antigen. STEAP gene is used in gene  
 CC therapy. Inhibiting the development or progression of a cancer (eg.  
 CC prostate, colon, bladder, lung, ovarian and pancreatic) expressing STEAP  
 CC or inhibiting growth or killing cells expressing STEAP in a patient;  
 CC comprises administering a vaccine composition to the patient. Treating  
 CC a patient with a cancer that expresses STEAP, or inhibiting growth or  
 CC killing cells expressing STEAP, comprises administering to the patient a  
 CC vector encoding single chain monoclonal antibody that comprises the  
 CC variable domains of the heavy and light chains of the monoclonal antibody  
 CC that specifically binds to STEAP, such that the vector delivers the  
 CC single chain monoclonal antibody coding sequence to the cancer cells and  
 CC the encoded single chain monoclonal antibody is expressed  
 CC intracellularly. The present sequence is an immunogenic peptide of  
 CC STEAP-1. STEAP-1 gene is located on chromosome 7p22.3. This peptide is  
 CC used to immunise sheep for the generation of sheep polyclonal antibodies  
 CC towards the amino-terminus of anti-STEAP-1.  
 CC  
 XX  
 SQ Sequence 15 AA;  
 Query Match 100.0%; Score 86; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKMKPRRLLEDDYL 15  
 DB 1 WKMKPRRLLEDDYL 15

RESULT 3  
 ID AAY12305 standard; Protein; 66 AA.  
 XX  
 AC AAY12305;  
 XX  
 DT 17-JUN-1999 (first entry)  
 XX  
 DE Human 5' EST secreted protein SEQ ID NO:336.  
 XX  
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW

KW chromolytic; anti-inflammatory; tumour inhibition.  
 XX  
 OS Homo sapiens.  
 XX  
 XX W09906548-A2.  
 XX  
 XX 11-FEB-1999.  
 XX  
 XX 31-JUL-1998; 98WO-IB01222.  
 XX  
 XX 01-AUG-1997; 97US-0905135.  
 XX  
 XX (GENSET).  
 XX  
 XX Ductert A, Dumas Milne Edwards J, Lacroix B;  
 PI  
 DR WPI; 1999-153778/13.  
 DR N-PSDB; AAX41138.  
 XX  
 XX New nucleic acids encoding human secreted proteins - obtained from  
 PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,  
 PT kidney, lung, umbilical cord, placenta and colon tissue  
 XX  
 XX Claim 27; Page 677-678; 824pp; English.  
 PS  
 CC AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins, and encode the proteins given in AAY12261 to  
 CC AAY12514, respectively. The proteins given represent the signal peptide  
 CC and an N-terminal fragment of a secreted protein. The nucleic acid  
 CC sequences can be used for producing secreted human gene products. They  
 CC can also be used to develop products for diagnosis and therapy. The  
 CC proteins obtained may have cytokine activity, cell  
 CC proliferation/differentiation activity, haematopoiesis regulating  
 CC activity, tissue growth regulating activity, reproductive hormone  
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and  
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, tumour inhibition activity or other activities. The products  
 CC can be used in forensic, gene therapy and chromosome mapping procedures.  
 CC The sequences can also be used for obtaining corresponding promoter  
 CC sequences. The nucleic acids encoding the signal peptide can be used for  
 CC directing extracellular secretion of a polypeptide or the insertion of a  
 CC polypeptide into a membrane, or importing a polypeptide into a cell.  
 CC  
 CC Sequence 66 AA;  
 SQ  
 Query Match 100.0%; Score 86; DB 20; Length 66;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 WKMKPRRNLEDDYL 15  
 Db 14 WKMKPRRNLEDDYL 28  
 RESULT 4  
 ID AAY12304 standard; Protein; 109 AA.  
 XX  
 XX AAY12304;  
 AC  
 XX  
 XX 17-JUN-1999 (first entry)  
 DT  
 XX  
 XX Human 5' EST secreted protein SEQ ID NO:335.  
 DE  
 XX  
 XX Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; anti-inflammatory; tumour inhibition.  
 KW  
 XX  
 XX Homo sapiens.  
 OS  
 XX

FN W09906548-A2.  
 XX  
 XX 11-FEB-1999.  
 XX  
 XX 31-JUL-1998; 98WO-IB01222.  
 XX  
 XX 01-AUG-1997; 97US-0905135.  
 XX  
 XX (GENSET).  
 XX  
 XX Ductert A, Dumas Milne Edwards J, Lacroix B;  
 PI  
 DR WPI; 1999-153778/13.  
 DR N-PSDB; AAX41137.  
 XX  
 XX New nucleic acids encoding human secreted proteins - obtained from  
 PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,  
 PT kidney, lung, umbilical cord, placenta and colon tissue  
 XX  
 XX Claim 27; Page 677; 824pp; English.  
 PS  
 CC AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins, and encode the proteins given in AAY12261 to  
 CC AAY12514, respectively. The proteins given represent the signal peptide  
 CC and an N-terminal fragment of a secreted protein. The nucleic acid  
 CC sequences can be used for producing secreted human gene products. They  
 CC can also be used to develop products for diagnosis and therapy. The  
 CC proteins obtained may have cytokine activity, cell  
 CC proliferation/differentiation activity, haematopoiesis regulating  
 CC activity, tissue growth regulating activity, reproductive hormone  
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and  
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, tumour inhibition activity or other activities. The products  
 CC can be used in forensic, gene therapy and chromosome mapping procedures.  
 CC The sequences can also be used for obtaining corresponding promoter  
 CC sequences. The nucleic acids encoding the signal peptide can be used for  
 CC directing extracellular secretion of a polypeptide or the insertion of a  
 CC polypeptide into a membrane, or importing a polypeptide into a cell.  
 CC  
 CC Sequence 109 AA;  
 SQ  
 Query Match 100.0%; Score 86; DB 20; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 3.8e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 WKMKPRRNLEDDYL 15  
 Db 14 WKMKPRRNLEDDYL 28  
 RESULT 5  
 ID AAY58194 standard; Protein; 339 AA.  
 XX  
 XX AAY58194;  
 AC  
 XX  
 XX 14-MAR-2000 (first entry)  
 DT  
 XX  
 XX Human STRAP-1 protein.  
 DE  
 XX  
 XX Serpentine transmembrane antigen of the prostate; STRAP-1; prostate;  
 KW transmembrane domain; type IIIa membrane protein; expression; cancer;  
 KW prostate cancer; bladder cancer; colon cancer; pancreatic cancer;  
 KW ovarian cancer; tumour antigen; immunisation; immune response;  
 KW cellular; humoral; anticancer vaccine; antibody; detection; diagnosis;  
 KW prognosis; monitoring; susceptibility; therapeutic inhibitor;  
 KW drug targeting; recombinant protein.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH 1..69  
 FT Region /note= "Intracellular region 1"  
 FT

FT	Domain	70..91	/note= "Transmembrane domain 1"
FT	Region	92..113	/note= "Extracellular region 1 (AAVS8198) "
FT	Domain	114..136	/note= "Transmembrane domain 2"
FT	Region	137..162	/note= "Intracellular region 2"
FT	Domain	163..184	/note= "Transmembrane domain 3"
FT	Region	185..218	/note= "Extracellular region 2 (AAVS8199) "
FT	Domain	219..241	/note= "Transmembrane domain 4"
FT	Region	242..253	/note= "Intracellular region 3"
FT	Domain	252..276	/note= "Transmembrane domain 5"
FT	Region	277..291	/note= "Extracellular region 3 (AAVS8200) "
FT	Domain	292..313	/note= "Transmembrane domain 6"
FT	Region	314..339	/note= "Intracellular region 4"
XX			
PN	W09962941.A2.		
XX			
PD	09-DEC-1999.		
XX			
PF	01-JUN-1999;	99WO-US12157.	
PR	01-JUN-1998;	98US-0087520.	
XX	30-JUN-1998;	98US-0091183.	
XX			
PA	(UROG-) UROGENESYS INC.		
PA	(AFAP-) AFAP D E.		
PA	(HUBE/) HUBERT R. S.		
PA	(LEON/) LEONG K.		
PA	(RAIT/) RAITANO A B.		
PA	(SAFE/) SAFEFRAN D C.		
PI	Afer DE, Hubert RS, Leong K, Raitano AB, Safiran DC;		
DR	WPI; 2000-072832/06.		
DR	N-PSDB; AA249395, AA249396.		
XX			
PT	Novel proteins useful as diagnostic markers and therapeutic targets,		
PT	particularly for prostatic cancer		
XX			
PS	Claim 1; Fig 1A; 83p; English.		
XX			
CC	This sequence represents a novel human protein, STRAP-1 (serpentine		
CC	transmembrane antigen of the prostate). STRAP-1 is the prototype		
CC	member of the STRAP family of proteins (AAVS8194-Y88197) which		
CC	exhibit a high degree of structural conservation, but which show		
CC	no significant structural homology to known human proteins. The STRAP-1		
CC	gene has been localised to chromosome 7p22. STRAP-1 is thought to be a		
CC	type IIIa membrane protein and is expressed predominantly in prostate		
CC	cells in normal human tissues. Structurally, STRAP-1 is a 339 amino		
CC	acid protein characterised by six transmembrane domains and		
CC	intracellular N- and C-termini, suggesting that it folds in a		
CC	"serpentine" manner into three extracellular and two intracellular loops		
CC	throughout all stages of prostate cancer. STRAP-1 mRNA and/or protein is		
CC	also overexpressed in certain other cancers, including bladder, colon,		
CC	pancreatic and ovarian cancer. The function of the STRAP proteins is not		
CC	known. They may be ion channels (from the presence of six transmembrane		
CC	domains), a feature which is shared by certain ion channels) or		
CC	gap-junction proteins (from immunohistochemical staining). STRAP-1 and		
CC	STRAP-2 are cell-surface tumour antigens. Immunisation with a STRAP		
CC	protein induces cellular and humoral immune responses against		
CC	STRAP-expressing cells. STRAP proteins may be used to identify		
CC	specific-binding agents, to produce anticancer vaccines and to generate		

CC	specific antibodies. The antibodies may be used for detection, prognosis,
CC	and monitoring of cancers (or susceptibility to cancer), as therapeutic
CC	inhibitors or to target therapeutic agents to their site of action. STRAP
CC	nucleic acids may be used for recombinant protein production, as
CC	diagnostic and prognostic reagents, for identifying STRAP-expressing
CC	cells for screening inhibitors of STRAP expression and for therapeutic
CC	modulation/inhibition of STRAP expression. Since high levels of STRAP
CC	proteins are exposed on the cell surface, they are easily targeted by
CC	systemically administered agents, and because they are expressed mainly
CC	on prostatic epithelial cells, agents targeted to them should have
CC	minimal side effects on other tissues.
XX	
XX	
SO	Sequence     339 AA;
OY	Query Match                    100.0%; Score 86; DB 21; Length 339; Best Local Similarity       100.0%; Pred. No. 1.3e-06; Matches    15; Conservative    0; Mismatches    0; Indels      0; Gaps       0;
Dd	1 WKMKPRRLLEDDYL 15       14 WKMKPRRLLEDDYL 28
RESULT 6	
AU6S9927	AU69927 standard; Protein: 339 AA.
ID	AU69927
AC	AU69927;
XX	
DT	30-JAN-2002 (first entry)
XX	
DE	Human prostate cDNA encoded protein #72.
XX	
KW	Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.
XX	
OS	Homo sapiens.
XX	
PN	WO200173032-A2.
XX	
PD	04-OCT-2001.
XX	
PE	27-MAR-2001; 2001WO-US09919.
XX	
FR	27-MAR-2000; 2000US-0536857.
PR	09-MAY-2000; 2000US-0568100.
PR	12-MAY-2000; 2000US-0570737.
PR	13-JUN-2000; 2000US-0593793.
PR	27-JUN-2000; 2000US-0605783.
PR	10-AUG-2000; 2000US-0636215.
PR	29-AUG-2000; 2000US-0651236.
PR	06-SEP-2000; 2000US-0657279.
PR	02-OCT-2000; 2000US-0679426.
PR	10-OCT-2000; 2000US-0685166.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Xu J, Dillon DC, Mitcham JT, Harlocker SL, Jiang Y, Kalos MD, Fanger GR, Retter MW, Scolk JA, Day CH, Vedvick TS, Carter D; Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA; WPI; 2001-639232/73. DR N-PsDB; AAS64160.
PT	New human prostate-specific polypeptides and polynucleotides useful for the diagnosis and treatment of cancer, especially prostate cancer -
PS	Claim 2; Page 549; 579pp; English.
XX	
XX	The invention relates to isolated prostate-specific polynucleotides, polypeptides, fusion proteins of the polypeptides, antibodies raised against the polypeptides (or antigenic epitopes derived from them) and antigen-presenting cells expressing the polypeptides. The antibodies are useful for detecting the presence of

CC cancer, especially prostate cancer. The polypeptides, polynucleotides and  
CC the antigen-presenting cells are useful for stimulating and/or expanding  
CC T cells specific for a tumour protein, and for inhibiting the development  
CC of cancer especially prostate cancer. Compositions comprising the  
CC polynucleotide and/or polypeptide are useful for stimulating an immune  
CC response, and for treating cancer. The oligonucleotide is useful for  
CC detecting cancer. The present sequence is a prostate specific  
CC polypeptide of the invention.  
SQ Sequence 339 AA;  
Query Match 100.0%; Score 86; DB 22; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 WKMKPRRNLEEDDYL 15  
DB 14 WKMKPRRNLEEDDYL 28  
RESULT 7  
AAM78845  
ID AAM78845 standard; Protein; 339 AA.  
XX  
AC AAM78845;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human protein SEQ ID NO 1507.  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
XX tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorder; arthritis; inflammation.  
XX Homo sapiens.  
XX  
XX WO200157190-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 05-FEB-2001; 2001WO-US04098.  
XX  
XX 03-FEB-2000; 2000US-0496914.  
XX 27-APR-2000; 2000US-0560875.  
XX 20-JUN-2000; 2000US-0598075.  
XX 19-JUL-2000; 2000US-0620325.  
XX 01-SEP-2000; 2000US-0654936.  
XX 15-SEP-2000; 2000US-0663561.  
XX 20-OCT-2000; 2000US-0693325.  
XX 30-NOV-2000; 2000US-0728422.  
XX  
XX (HUSB-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R,  
XX WPI; 2001-476283/51.  
XX N-PSDB; AAK51978.  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities,  
XX useful in diagnosis and gene therapy -  
XX  
XX Claim 20; Page 3800-3801; 6221PP; English.  
XX  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
XX encoded polypeptides (AAM78323-AAW60302) that exhibit activity elating to  
XX cytokine, cell proliferation or cell differentiation or which may induce  
XX production of other cytokines in other cell populations. The  
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or  
XX peptide therapy. The polypeptides have various cytokine-like activities,  
XX e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activity/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80920) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
SQ Sequence 339 AA;  
Query Match 100.0%; Score 86; DB 22; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 WKMKPRRNLEEDDYL 15  
DB 14 WKMKPRRNLEEDDYL 28  
RESULT 8  
AAM01282  
ID AAM01282 standard; Protein; 339 AA.  
XX  
AC AAM01282;  
XX  
DT 04-OCT-2001 (first entry)  
XX  
DE F789P amino acid sequence.  
XX  
XX Human; prostate cancer; prostate-specific; diagnosis; vaccine;  
XX cytostatic; gene therapy; metastasis.  
XX  
XX Homo sapiens.  
XX  
XX WO200151633-A2.  
XX  
XX 19-JUL-2001.  
XX  
XX 16-JAN-2001; 2001WO-US01574.  
XX  
XX 14-JAN-2000; 2000US-0483672.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Xu J, Dillon DC, Mitcham JL, Harlocker SH, Jiang Y, Reed SG;  
XX Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YMW;  
XX Wang A, Meagner MJ;  
XX WPI; 2001-425873/45.  
XX  
XX New polynucleotide encoding a prostate-specific protein, for  
XX diagnosing, monitoring and treating prostate cancer in a patient and  
XX for use in vaccines -  
XX  
XX Claim 2; Page 510-512; 543PP; English.  
XX  
XX The present invention describes polynucleotide sequences (I) which encode  
XX prostate-specific proteins (II). (I) and (II) have cytosratic activity,  
XX and can be used in vaccine production and gene therapy. (I), (II),  
XX antibodies to (II), fusion proteins comprising (II), and isolated  
XX T cells prepared using (I) or (II) are used treat cancer in a patient.  
XX (I) and the antibodies are also used in the detection of cancer in a  
XX patient. The cancer that is diagnosed or treated is particularly  
XX prostate cancer. (I) and (II) can be used in vaccines. The antibodies or  
XX (I) can be used for monitoring the progression of cancer in a patient.  
XX (I) and (II) can also be used to improve diagnostic and therapeutic  
XX methods for prostate cancer. They can indicate the level of metastasis  
XX as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to  
XX AAM01318 represent polynucleotide and amino acid sequences used in the  
XX exemplification of the present invention.

Query Match 100.0%; Score 86; DB 22; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKMKPRRLLEDDYL 15  
DB 14 WKMKPRRLLEDDYL 28

## RESULT 9

ABG61813  
ID ABG61813 standard; Protein; 339 AA.

AC ABG61813;

DT 15-AUG-2002 (first entry)

DE Prostate cancer-associated protein #14.

XX Prostate cancer; prostate tumour tissue; human; mammal; cytosstatic.

XX Mammalia.

PN W0200230268-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001MO-US32045.

PR 13-OCT-2000; 2000US-0687576.

PR 08-DEC-2000; 2000US-0733288.

PR 08-DEC-2000; 2000US-0733742.

PR 24-JAN-2001; 2001US-263957P.

PR 16-MAR-2001; 2001US-276791P.

PR 16-MAR-2001; 2001US-276888P.

PR 06-APR-2001; 2001US-281922P.

PR 24-APR-2001; 2001US-286214P.

PR 30-APR-2001; 2001US-0847046.

PR 04-MAY-2001; 2001US-288589P.

XX (E0SB-) EOS BIOTECHNOLOGY INC.

PA Gish KC, Mack DH, Wilson KE, Afari D, Hevezl P;

PI WPI; 2002-471335/50.

DR N-PSDB; ABR92128.

XX Detecting a prostate cancer-associated transcript in a cell in a

PT patient, useful for diagnosing prostate cancer (PC) or screening

PT modulators of PC, by determining if prostate cancer-associated genes

XX are expressed in a prostate tissue

XX Claim 27; Page 312; 436pp; English.

XX The present invention relates to methods of detecting a prostate

CC cancer-associated transcript in a cell from a patient. The method

CC comprises contacting a biological sample from the patient with

CC prostate cancer-associated polynucleotides (designated PC genes) that

CC selectively hybridize to a sequence that is at least 80% identical

CC to them. The prostate cancer-associated polynucleotide sequences

CC are differentially expressed in prostate tumour tissue or in

CC prostate cancer and are derived from the tissues of various

CC organisms such as humans or other mammals (e.g. mice, sheep and dogs).

CC The methods of the invention are useful for diagnosing and treating

CC prostate cancer in mammals. The prostate cancer-associated genes are

CC useful for diagnosing or treating prostate cancer, as well as for

CC identifying modulators of prostate cancer or agents that inhibit

CC prostate cancer. The nucleic acid sequences are particularly useful

CC in gene therapy, as a vaccine or in antisense applications.

CC ABG61800-ABG61944 represent prostate cancer-associated proteins.

XX Sequence 339 AA;

Query Match 100.0%; Score 86; DB 23; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKMKPRRLLEDDYL 15  
DB 14 WKMKPRRLLEDDYL 28

## RESULT 10

ABB95387  
ID ABB95387 standard; Protein; 339 AA.

AC ABB95387;

DT 19-JUL-2002 (first entry)

DE Human P789P protein SEQ ID NO 879.

XX Human; cancer; prostate cancer; vaccine; cytosstatic; immunostimulant;

XX gene therapy.

XX Homo sapiens.

PN US200202248-A1.

PD 21-FEB-2002.

PF 12-JAN-2001; 2001US-0759143.

PR 25-FEB-1997; 97US-0806099.

PR 01-AUG-1997; 97US-0904804.

PR 09-FEB-1998; 98US-0020956.

PR 25-FEB-1998; 98US-0030607.

PR 14-JUL-1998; 98US-0115453.

PR 23-SEP-1998; 98US-0159812.

PR 15-JAN-1999; 99US-0232149.

PR 09-APR-1999; 99US-0288946.

PR 13-JUL-1999; 99US-0352616.

PR 12-NOV-1999; 99US-0439313.

PR 18-NOV-1999; 99US-0443686.

PR 14-JAN-2000; 2000US-0483672.

PR 27-MAR-2000; 2000US-0536857.

PR 09-MAY-2000; 2000US-0568100.

PR 12-MAY-2000; 2000US-0570737.

PR 13-JUN-2000; 2000US-0593793.

PR 27-JUN-2000; 2000US-0605783.

PR 10-AUG-2000; 2000US-0636215.

PR 29-AUG-2000; 2000US-0651236.

PR 06-SEP-2000; 2000US-0657279.

PR 02-OCT-2000; 2000US-0679426.

PR 10-OCT-2000; 2000US-0685166.

XX (XUJ/) XU J.

PA (DILL/) DILLON D C.

PA (MITC/) MITCHAM J L.

PA (HARL/) HARLOCKER S L.

PA (JIAN/) JIANG Y.

PA (KALO/) KALOS M D.

PA (FANG/) FANGER G R.

PA (RETT/) RETTER M W.

PA (STOL/) STOLK J A.

PA (DAYC/) DAY C H.

PA (VEDV/) VEDVICK T S.

PA (CART/) CARTER D.

PA (LISK/) LI S X.

PA (WANG/) WANG A.

PA (SKEI/) SKEIKY Y A W.

PA (HEPL/) HEPLER M T.

PA (HEND/) HENDERSON R A.

XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;

PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;



PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;  
XX WPI; 2002-255649/30.  
XX  
XX New prostate-specific polynucleotides for diagnosing and treating  
PT disease, in particular prostate cancer, and as markers for the  
XX progression of cancer  
XX  
PS Claim 2; SEQ ID NO 879; 87pp; English.  
XX  
CC The present invention provides prostate-specific coding sequences and  
CC their encoded proteins. These can be used in the diagnosis and treatment  
CC of cancers, particularly prostate cancer. The present sequence is a  
CC protein described in the invention.  
XX  
SQ Sequence 339 AA;  
  
Query Match 100.0%; Score 86; DB 23; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 WKMKPRRNLEDDYL 15  
Db 14 WKMKPRRNLEDDYL 28  
|||||  
  
RESULT 11  
AAM79829  
ID AAM79829 standard; Protein; 374 AA.  
XX  
XX AAM79829;  
AC  
XX  
XX 06-NOV-2001 (first entry)  
DT  
XX  
XX Human protein SEQ ID NO 3475.  
DE  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200157190-A2.  
PN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 05-FEB-2001; 2001WO-US04098.  
PF  
XX  
XX 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX  
XX (HYSB-) HYSBQ INC.  
PA  
XX  
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;  
XX  
XX WPI; 2001-476283/51.  
DR N-PSDB; AAK52962.  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
XX  
XX Claim 20; Page 350; 6221pp; English.  
PS  
XX  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the

CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
XX  
SQ Sequence 374 AA;  
  
Query Match 100.0%; Score 86; DB 22; Length 374;  
Best Local Similarity 100.0%; Pred. No. 1.4e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 WKMKPRRNLEDDYL 15  
Db 49 WKMKPRRNLEDDYL 63  
|||||  
  
RESULT 12  
AAE02780  
ID AAE02780 standard; Protein; 375 AA.  
XX  
XX AAE02780;  
AC  
XX  
XX 06-AUG-2001 (first entry)  
DT  
XX  
XX Human six transmembrane epithelial antigen of prostate (STREP)-1 protein.  
DE  
XX  
XX Human; cytostatic; antiproliferative; vaccine; gene therapy;  
KW six transmembrane epithelial antigen of the prostate-1; STREP-1;  
KW chromosome 7p22.3; cancer; prostate; colon; bladder; lung; ovarian;  
KW pancreatic.  
XX  
XX Homo sapiens.  
OS  
XX  
XX  
FH Key Location/Qualifiers  
XX  
XX Region 14..28  
FT /label= Immunogenic\_peptide #1  
XX  
XX Domain 70..91  
FT /label= Transmembrane\_domain #1  
XX  
XX Region 86..94  
FT /label= HLA-A2\_binding\_peptide #2  
XX  
XX Domain 114..136  
FT /label= Transmembrane\_domain #2  
XX  
XX Region 158..166  
FT /label= HLA-A2\_binding\_peptide #5  
XX  
XX Domain 163..184  
FT /label= Transmembrane\_domain #3  
XX  
XX Region 165..173  
FT /label= HLA-A2\_binding\_peptide #1  
XX  
XX Domain 219..241  
FT /label= Transmembrane\_domain #4  
XX  
XX Domain 254..276  
FT /label= Transmembrane\_domain #5  
XX  
XX Region 262..270  
FT /label= HLA-A2\_binding\_peptide #3  
XX  
XX Domain 292..313  
FT /label= Transmembrane\_domain #6  
XX  
XX Region 302..310  
FT /label= HLA-A2\_binding\_peptide #4  
XX  
XX Misc-difference 339..340  
FT /note= "Encoded by TTGRAAAT"  
XX  
XX  
XX WO200140276-A2.  
PN  
XX  
XX 07-JUN-2001.  
PD

```

XX 06-DEC-2000; 2000MO-US33040.
PF
XX
XX 06-DEC-1999; 99US-0455486.
PR
XX
XX (UROG-) UROGENESYS INC.
PA
XX
XX Afar DEH, Hubert RS, Raitano AB, Saffran DC, Mitchell SC, Faris M,
PI Jakobovits A;
PI
XX
XX MPI: 2001-367804/38.
DR
XX N-PSDB; AAD07067.
XX
XX New STEAP (six transmembrane epithelial antigen of the prostate)
PT protein, expressed in human cancers, useful for detecting and treating
PT cancer -
XX
XX Example 2; Fig 1A-1B; 187pp; English.
PS
XX
XX The present sequence is human six transmembrane epithelial antigen of
CC the prostate (STEAP)-1 protein of clone 10. STEAP-1 is a member of cell
CC surface serpentine transmembrane antigens. STEAP-1 gene is located on
CC chromosome 7p22.3 and is used in gene therapy. Inhibiting the development
CC or progression of a cancer (eg. prostate, colon, bladder, lung, ovarian
CC and pancreatic) expressing STEAP or inhibiting growth or killing cells
CC expressing STEAP in a patient, comprises administering a vaccine
CC composition to the patient. Treating a patient with a cancer that
CC expresses STEAP, or inhibiting growth or killing cells expressing STEAP,
CC comprises administering to the patient a vector encoding single chain
CC monoclonal antibody that comprises the variable domains of the heavy and
CC light chains of the monoclonal antibody that specifically binds to STEAP,
CC such that the vector delivers the single chain monoclonal antibody coding
CC sequence to the cancer cells and the encoded single chain monoclonal
CC antibody is expressed intracellularly.
CC Note: The present sequence is also shown in sequence listing of the
CC specification, but it lacks amino acid residues at its N-terminal end.
XX
XX Sequence 375 AA;
SQ
XX
XX
XX Query Match 100.0%; Score 86; DB 22; Length 375;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-06;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 WKMKPRRLLEDDYL 15
XX |||||
XX 14 WKMKPRRLLEDDYL 28
XX
XX
XX RESULT 13
XX AAY11840
XX ID AAY11840 standard; Protein; 95 AA.
XX
XX AAY11840;
XX
XX 18-JUN-1999 (first entry)
XX
XX Human 5' EST secreted protein SEQ ID No: 440.
XX
XX Human; secreted protein; EST; expressed sequence tag; diagnosis;
XX forensic; gene therapy; chromosome mapping; signal peptide; prostate;
XX upstream regulatory sequence; cytokine activity; cell proliferation;
XX differentiation; haematopoiesis regulation; tissue growth regulation;
XX reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
XX thrombolytic; anti-inflammatory; tumour inhibition.
XX
XX Homo sapiens.
XX
XX WO9906550-A2.
XX
XX 11-FEB-1999.
XX
XX 31-JUL-1998; 98MO-IB01232.
XX

```

```

PR 01-AUG-1997; 97US-0905144.
XX
XX (GEST ) GENSET.
PA
XX
XX Duclert A, Dumas Milne Edwards J, Lacroix B;
PI
XX
XX MPI: 1999-153780/13.
DR
XX N-PSDB; AAX40562.
XX
XX New isolated prostate-derived nucleic acids - used to develop
PT products which may have cytokine, immune regulatory, haematopoiesis
PT regulating, anti-inflammatory or tumour inhibition activity
XX
XX Claim 34; Page 577; 675pp; English.
PS
XX
XX AAX40438 to AAX40715 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins expressed in prostate, and encode the proteins
CC given in AAY11716 to AAY11993 respectively. The proteins given represent
CC the signal peptide and an N-terminal fragment of a secreted protein. The
CC nucleic acid sequences can be used for producing secreted human gene
CC products. They can also be used to develop products for diagnosis and
CC therapy. The proteins obtained may have cytokine activity, cell
CC proliferation and differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptides can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
XX
XX Sequence 95 AA;
SQ
XX
XX
XX Query Match 81.4%; Score 70; DB 20; Length 95;
XX Best Local Similarity 100.0%; Pred. No. 0.0002;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 3 MKPRRLLEDDYL 15
XX |||||
XX 1 MKPRRLLEDDYL 13
XX
XX
XX RESULT 14
XX AAW86309
XX ID AAW86309 standard; Protein; 339 AA.
XX
XX AAW86309;
XX
XX 01-MAR-1999 (first entry)
XX
XX Kidney injury associated molecule HM018 protein.
XX
XX Kidney injury associated molecule; kidney injury related molecule;
XX kidney injury associated molecule; kidney injury related molecule;
XX kidney injury associated molecule; kidney injury related molecule;
XX acute renal failure; acute nephritis; tumour.
XX
XX Rattus sp.
XX
XX WO9853071-A1.
XX
XX 26-NOV-1998.
XX
XX 22-MAY-1998; 98MO-US10547.
XX
XX 23-MAY-1997; 97US-0047491.
XX
XX 23-MAY-1997; 97US-0047490.
XX
XX (BIO ) BIOGEN INC.
XX
XX Cate RL, Hession CA, Sanicola-Nadel M, Wei H;
XX

```

DR WPI: 1999-045312/04.  
 DR N-PSDB; AAV80586.  
 XX  
 PT Kidney injury-associated molecule, KIM, polypeptides - upregulated  
 PT in injured or regenerating tissues, useful to promote tissue growth  
 PT and regeneration, especially to treat renal conditions  
 XX  
 PS Claim 17; Page 57-58; 213pp; English.  
 XX  
 CC The present sequence represents a kidney injury associated molecule  
 CC (KIM) protein. KIM proteins can be administered therapeutically  
 CC by expressing KIM encoding polynucleotides, to promote growth and/or  
 CC survival of damaged tissue (e.g. renal tissue), since the KIM proteins  
 CC are upregulated in injured or regenerating (especially renal) tissues.  
 CC KIM fusion proteins, conjugates, antibodies and vectors can also be used  
 CC therapeutically, e.g. these or the KIM proteins may be included with an  
 CC acceptable carrier in pharmaceutical compositions, useful for therapy/  
 CC prophylaxis of conditions associated with dysfunction/dysregulation of  
 CC KIM genes or proteins, especially renal diseases or impairments of renal  
 CC function in humans (e.g. acute renal failure, acute nephritis). The  
 CC polynucleotides can be used to produce antisense sequences which, when  
 CC internalised into cells, can disrupt expression of a cellular KIM gene,  
 CC also useful in therapy (e.g. to block the growth of tumours dependent on  
 CC KIM for growth) or compositions. The proteins and polynucleotides are  
 CC useful diagnostically e.g. to detect and quantify renal injury/disease  
 CC (indicative of increased risk, or presence of, renal injury or impaired  
 CC function), or abnormal responses to tissue injury (indicative of  
 CC increased risk, or presence of, an autoimmune response or abnormal  
 CC tissue growth arising from/affecting renal tissue). The proteins can  
 CC also be used to locate KIM-producing cells (especially specific loci,  
 CC e.g. tissue masses abnormally producing/expressing KIM such as tumours  
 CC arising from/affecting renal tissue), by contacting cells with an  
 CC imageable KIM-binding reagent and imaging reagent accumulation.  
 CC  
 SQ Sequence 339 AA;  
 Query Match 73.3%; Score 63; DB 20; Length 339;  
 Best Local Similarity 71.4%; Pred. No. 0.012;  
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 WKMKPRRNLEDDY 14  
 |||||:||||  
 Db 14 WKMKPKGNLEDDSY 27  
 |||||:||||  
 RESULT 15  
 ABG25517  
 ID ABG25517 standard; Protein; 72 AA.  
 XX  
 AC ABG25517;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #25508.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX

DR WPI: 2001-639362/73.  
 DR N-PSDB; AAS89704.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 55876; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (II) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 72 AA;  
 Query Match 52.3%; Score 45; DB 22; Length 72;  
 Best Local Similarity 63.6%; Pred. No. 3.1;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 WKMKPRRNLEE 11  
 |||||:||||  
 Db 39 WKMKPQNTLPE 49  
 |||||:||||

Search completed: March 26, 2003, 16:48:00  
 Job time : 33.1279 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: March 26, 2003, 16:49:57 ; Search time 10.1163 Seconds  
(without alignments)  
87.073 Million cell updates/sec

Title: US-10-010-667A-2\_COPY\_14\_28

Perfect score: 86  
Sequence: 1 WKMKRRRLREDDYL 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 237916 seqs, 58723674 residues

Total number of hits satisfying chosen parameters: 237916

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	86	100.0	339	9	US-10-012-896-879
2	86	100.0	339	9	US-09-802-520-11
3	86	100.0	339	9	US-09-895-793-879
4	86	100.0	339	9	US-09-895-814-879
5	86	100.0	339	9	US-10-011-095-2
6	86	100.0	339	9	US-10-010-667A-2
7	86	100.0	339	10	US-09-759-143-879
8	86	100.0	339	10	US-09-780-669-879
9	86	100.0	339	10	US-09-822-827-879
10	86	100.0	339	10	US-09-801-574-32
11	42	48.8	1019	10	US-09-801-574-76
12	39	45.3	262	9	US-09-813-453A-45
13	38	44.2	404	9	US-09-486-734A-12
14	38	44.2	539	10	US-09-815-242-10497
15	38	44.2	968	10	US-09-753-008-7
16	37.5	43.6	405	9	US-10-102-806-675
17	37.5	43.6	803	9	US-09-968-436B-2
18	37.5	43.6	803	10	US-09-759-010-7
19	37	43.0	46	10	US-09-864-761-34052

20	37	43.0	253	9	US-10-078-650-18	Sequence 18, Appl
21	37	43.0	340	10	US-09-925-302-448	Sequence 448, Appl
22	37	43.0	410	9	US-10-078-650-14	Sequence 14, Appl
23	37	43.0	482	9	US-09-746-783-66	Sequence 66, Appl
24	37	43.0	522	10	US-09-817-764-3	Sequence 3, Appl
25	37	43.0	577	9	US-09-893-519A-61	Sequence 61, Appl
26	37	43.0	605	9	US-10-028-072-160	Sequence 160, Appl
27	37	43.0	605	9	US-10-121-049-160	Sequence 160, Appl
28	37	43.0	605	9	US-10-123-904-160	Sequence 160, Appl
29	37	43.0	605	9	US-10-140-470-160	Sequence 160, Appl
30	37	43.0	605	9	US-09-833-129-37	Sequence 37, Appl
31	37	43.0	605	9	US-10-175-746-160	Sequence 160, Appl
32	37	43.0	605	9	US-10-176-918-160	Sequence 160, Appl
33	37	43.0	605	9	US-10-176-921-160	Sequence 160, Appl
34	37	43.0	605	9	US-10-137-865-160	Sequence 160, Appl
35	37	43.0	605	9	US-10-140-474-160	Sequence 160, Appl
36	37	43.0	605	9	US-10-143-431-160	Sequence 160, Appl
37	37	43.0	605	9	US-10-143-114-160	Sequence 160, Appl
38	37	43.0	605	9	US-10-140-002-160	Sequence 160, Appl
39	37	43.0	605	9	US-10-142-419-160	Sequence 160, Appl
40	37	43.0	605	9	US-10-123-262-160	Sequence 160, Appl
41	37	43.0	605	9	US-10-142-423-160	Sequence 160, Appl
42	37	43.0	605	9	US-10-121-050-160	Sequence 160, Appl
43	37	43.0	605	9	US-10-141-755-160	Sequence 160, Appl
44	37	43.0	605	10	US-09-860-298-2	Sequence 2, Appl
45	37	43.0	635	9	US-10-101-464A-932	Sequence 932, Appl

#### ALIGNMENTS

RESULT 1  
US-10-012-896-879  
Sequence 879, Application US/10012896  
Publication No. US20020183251A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Dillon, David C.  
APPLICANT: Micham, Jennifer L.  
APPLICANT: Harlocker, Susan L.  
APPLICANT: Jiang, Yugu  
APPLICANT: Kalos, Michael D.  
APPLICANT: Retter, Marc W.  
APPLICANT: Stolk, John A.  
APPLICANT: Day, Craig H.  
APPLICANT: Vedrick, Thomas S.  
APPLICANT: Carter, Darrick  
APPLICANT: Li, Samuel X.  
APPLICANT: Wang, Aijun  
APPLICANT: Skeiky, Yaelir A.W.  
APPLICANT: Hepner, William T.  
APPLICANT: Henderson, Robert A.  
APPLICANT: Hural, John  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Houghton, Raymond L.  
APPLICANT: Vinals de Bassols, Carlota  
APPLICANT: Foy, Teresa  
APPLICANT: Fanger, Gary R.  
APPLICANT: Watanabe, Yoshinhiro  
APPLICANT: Meagher, Madeleine Joy  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
FILE REFERENCE: 210121.427C27  
CURRENT APPLICATION NUMBER: US/10/012,896  
CURRENT FILING DATE: 2001-12-10  
NUMBER OF SEQ ID NOS: 1011  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 879  
LENGTH: 339  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-012-896-879

Query Match 100.0%; Score 86; DB 9; Length 339;  
Best Local Similarity 100.0%; Pred. No. 9.5e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKMKPRRNLEDDYL 15  
|||||  
DB 14 WKMKPRRNLEDDYL 28

## RESULT 2

US-09-802-520-11  
; Sequence 11, Application US/09802520  
; Publication No. US20020187472A1  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Faris, Mary  
; APPLICANT: Chen, Hui-Mei  
; APPLICANT: Ison, Craig H.  
; TITLE OF INVENTION: STEAP-RELATED PROTEIN  
; FILE REFERENCE: PC-0037 US  
; CURRENT APPLICATION NUMBER: US/09/802,520  
; CURRENT FILING DATE: 2001-03-09  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PERL Program  
; SEQ ID NO 11  
; LENGTH: 339  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. US20020187472A1 96572948  
US-09-802-520-11

Query Match 100.0%; Score 86; DB 9; Length 339;  
Best Local Similarity 100.0%; Pred. No. 9.5e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKMKPRRNLEDDYL 15  
|||||  
DB 14 WKMKPRRNLEDDYL 28

## RESULT 3

US-09-895-793-879  
; Sequence 879, Application US/09895793  
; Publication No. US20020192763A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiaqun  
; APPLICANT: Dillon, David C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan L.  
; APPLICANT: Jiang, Yugu  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Stolck, John A.  
; APPLICANT: Day, Craig H.  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Carter, Darick  
; APPLICANT: Li, Samuel X.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Hepler, William T.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Hural, John  
; APPLICANT: McNeill, Patricia D.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Vinals de Basols, Carlotca  
; APPLICANT: Foy, Teresa  
; APPLICANT: Fanger, Gary R.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
; FILE REFERENCE: 210121.534C2  
; CURRENT APPLICATION NUMBER: US/09/895,793

; CURRENT FILING DATE: 2001-06-29  
; NUMBER OF SEQ ID NOS: 982  
; SOFTWARE: FaastSeq for Windows Version 3.0  
; SEQ ID NO 879  
; LENGTH: 339  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-895-793-879

QY 1 WKMKPRRNLEDDYL 15  
|||||  
DB 14 WKMKPRRNLEDDYL 28

Query Match 100.0%; Score 86; DB 9; Length 339;  
Best Local Similarity 100.0%; Pred. No. 9.5e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKMKPRRNLEDDYL 15  
|||||  
DB 14 WKMKPRRNLEDDYL 28

## RESULT 4

US-09-895-814-879  
; Sequence 879, Application US/09895814  
; Publication No. US20020193286A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiaqun  
; APPLICANT: Dillon, David C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan L.  
; APPLICANT: Jiang, Yugu  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Stolck, John A.  
; APPLICANT: Day, Craig H.  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Carter, Darick  
; APPLICANT: Li, Samuel X.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Hepler, William T.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Hural, John  
; APPLICANT: McNeill, Patricia D.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Vinals de Basols, Carlotca  
; APPLICANT: Foy, Teresa  
; APPLICANT: Fanger, Gary R.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
; FILE REFERENCE: 210121.427C26  
; CURRENT APPLICATION NUMBER: US/09/895,814  
; CURRENT FILING DATE: 2001-06-29  
; NUMBER OF SEQ ID NOS: 990  
; SOFTWARE: FaastSeq for Windows Version 3.0  
; SEQ ID NO 879  
; LENGTH: 339  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-895-814-879

Query Match 100.0%; Score 86; DB 9; Length 339;  
Best Local Similarity 100.0%; Pred. No. 9.5e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKMKPRRNLEDDYL 15  
|||||  
DB 14 WKMKPRRNLEDDYL 28

## RESULT 5

US-10-011-095-2  
; Sequence 2, Application US/10011095  
; Publication No. US20030045682A1  
; GENERAL INFORMATION:  
; APPLICANT: Afar, Daniel

```
APPLICANT: Hubert, Rene S.
APPLICANT: Leong, Kahan
APPLICANT: Raitano, Arthur B.
APPLICANT: Saffran, Douglas C.
APPLICANT: Mitchell, Steve Chappell
TITLE OF INVENTION: ANTIBODIES IMMUNOSPECIFIC FOR STEAP1 (AS AMENDED)
FILE REFERENCE: 511582001610
CURRENT FILING DATE: US/10/011,095
PRIOR FILING DATE: 2001-12-06
PRIOR APPLICATION NUMBER: 09/323,873
PRIOR FILING DATE: 1998-06-01
PRIOR APPLICATION NUMBER: 60/087,520
PRIOR FILING DATE: 1998-06-01
PRIOR APPLICATION NUMBER: 60/091,183
PRIOR FILING DATE: 1998-06-30
NUMBER OF SEQ ID NOS: 32
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 339
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: DNA
US-10-011-095-2
```

```
Query Match          100.0%; Score 86; DB 9; Length 339;
Best Local Similarity 100.0%; Pred. No. 9.5e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WKMKPRRNLEEDYDL 15
        |||
        14 WKMKPRRNLEEDYDL 28
```

```
RESULT 6
US-10-010-667a-2
Sequence 2, Application US/10010667A
Publication No. US20030055217A1
GENERAL INFORMATION:
APPLICANT: Afar, Daniel
APPLICANT: Hubert, Rene S.
APPLICANT: Leong, Kahan
APPLICANT: Raitano, Arthur B.
APPLICANT: Saffran, Douglas C.
APPLICANT: Mitchell, Steve Chappell
TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
EXPRESSED IN HUMAN CANCERS AND USES THEREOF
FILE REFERENCE: 511582001601
CURRENT FILING DATE: US/10/010,667A
CURRENT FILING DATE: 2001-12-06
PRIOR APPLICATION NUMBER: 09/323,873
PRIOR FILING DATE: 1998-06-01
PRIOR APPLICATION NUMBER: 60/087,520
PRIOR FILING DATE: 1998-06-01
PRIOR APPLICATION NUMBER: 60/091,183
PRIOR FILING DATE: 1998-06-30
NUMBER OF SEQ ID NOS: 32
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 339
TYPE: PRT
ORGANISM: Homo sapiens
US-10-010-667a-2
```

```
Query Match          100.0%; Score 86; DB 9; Length 339;
Best Local Similarity 100.0%; Pred. No. 9.5e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WKMKPRRNLEEDYDL 15
        |||
        14 WKMKPRRNLEEDYDL 28
```

```
RESULT 7
US-09-759-143-879
Sequence 879, Application US/09759143
Patent No. US20020022248A1
GENERAL INFORMATION:
APPLICANT: Xu, Jianshun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqun
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C23
CURRENT FILING DATE: US/09/759,143
CURRENT FILING DATE: 2001-01-12
NUMBER OF SEQ ID NOS: 934
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 879
LENGTH: 339
TYPE: PRT
ORGANISM: Homo sapiens
US-09-759-143-879
```

```
Query Match          100.0%; Score 86; DB 10; Length 339;
Best Local Similarity 100.0%; Pred. No. 9.5e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WKMKPRRNLEEDYDL 15
        |||
        14 WKMKPRRNLEEDYDL 28
```

```
RESULT 8
US-09-780-669-879
Sequence 879, Application US/09780669
Patent No. US20020051977A1
GENERAL INFORMATION:
APPLICANT: Xu, Jianshun
APPLICANT: Dillon, Davin C.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqun
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C24
CURRENT APPLICATION NUMBER: US/09/780,669
```

```

; CURRENT FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 943
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 879
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-780-669-879
```

```

Query Match          100.0%; Score 86; DB 10; Length 339;
Best Local Similarity 100.0%; Pred. No. 9,5e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 WKMKPRNLEDDYL 15
Db      14 WKMKPRNLEDDYL 28
```

```

RESULT 9
US-09-822-827-879
; Sequence 879, Application US/09822827
; Patent No. US20020081680A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.534C1
; CURRENT APPLICATION NUMBER: US/09/822,827
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 879
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-822-827-879
```

```

Query Match          100.0%; Score 86; DB 10; Length 339;
Best Local Similarity 100.0%; Pred. No. 9,5e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 WKMKPRNLEDDYL 15
Db      14 WKMKPRNLEDDYL 28
```

```

RESULT 10
US-09-801-574-32
; Sequence 32, Application US/09801574
; Patent No. US20020081592A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Peijiang Jeremy
; TITLE OF INVENTION: Reproduction-Specific Genes
; FILE REFERENCE: 0399.2007-002
; CURRENT APPLICATION NUMBER: US/09/801,574
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: 60/187,518
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: 60/261,557
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 32
; LENGTH: 1018
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-801-574-32
```

```

Query Match          48.8%; Score 42; DB 10; Length 1018;
Best Local Similarity 53.8%; Pred. No. 73;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
```

```

Qy      1 WKMKPRNLEDD 13
Db      646 WKLVKSNLEESD 658
```

```

RESULT 11
US-09-801-574-76
; Sequence 76, Application US/09801574
; Patent No. US20020081592A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Peijiang Jeremy
; TITLE OF INVENTION: Reproduction-Specific Genes
; FILE REFERENCE: 0399.2007-002
; CURRENT APPLICATION NUMBER: US/09/801,574
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: 60/187,518
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: 60/261,557
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 76
; LENGTH: 1019
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-801-574-76
```

```

Query Match          48.8%; Score 42; DB 10; Length 1019;
Best Local Similarity 53.8%; Pred. No. 74;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy      1 WKMKPRNLEED 13
Db      645 WKLVKSNLEESD 657
```

```

RESULT 12
US-09-813-453A-45
; Sequence 45, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OCG-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 45
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-09-813-453A-45
```

```

Query Match          45.3%; Score 39; DB 9; Length 262;
Best Local Similarity 42.9%; Pred. No. 53;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy      1 WKMKPRNLEEDY 14
Db      26 WRMETDRHKTEDY 39
```

```

RESULT 13
US-09-486-734A-12
; Sequence 12, Application US/09486734A
```



```

Query March          44.2%; Score 38; DB 10; Length 539;
Best Local Similarity 46.7%;
                          Pred. No. 1.7e+02;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY      1 WKMKPRRLNEEDDTL 15
      | : | | | : | | |
Db      22 WFKVKGDTINEEDDTL 36

```

```

RESULT 15
US-09-753-008-7
; Sequence 7, Application US/09753008
; Patent No. US20020061520A1
; GENERAL INFORMATION:
; APPLICANT: Stefan Somlo and Toshio Mochizuki
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE PKD2 GENE
; AND USES THEREOF
;
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE
; DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/753,008
; FILING DATE: 02-Jan-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/385,752
; FILING DATE: 1999-08-30
; APPLICATION NUMBER: 08/651,999
; FILING DATE: MAY 23, 1996
;
; ATTORNEY/AGENT INFORMATION:
; NAME: ELIZABETH A. BOGOSIAN
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 96700/395
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
;
; INFORMATION FOR SEQ ID NO: 7
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 968 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: <Unknown>
; DESCRIPTION: peptide
; HYPOTHEetical: NO
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 7
;
;
US-09-753-008-7

Query Match          44.2%; Score 38; DB 10; Length 968;
Best Local Similarity 46.2%; Pred. No. 3.2e+02;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0

QY 1 WKMPRLNEEDD 13
   |||:|:|:
Db 293 WKOPSNOTEDN 305

Search completed: March 26, 2003, 17:04:14
Job time : 11.1163 secs

```



GenCore version 5.1.4 ps 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 16:43:45 / Search time 15.6977 Seconds  
(without alignments)  
196.889 Million cell updates/sec

Title: US-10-010-667a-2\_COPY\_14\_28  
Perfect score: 86  
Sequence: 1 WKKKPRRLIEDDYL 15

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

SPTRMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp Unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	78	90.7	264	4	Q9Y6U5	Q9Y6U5 homo sapien
2	67.5	78.5	338	6	Q9GL50	Q9GL50 sus scrofa
3	63	73.3	339	11	Q9CWR7	Q9CWR7 mus musculu
4	63	73.3	339	11	Q924Z2	Q924Z2 mus musculu
5	63	73.3	339	11	Q924J9	Q924J9 mus musculu
6	45.5	52.9	237	3	Q9URAS	Q9URAS schizosacch
7	45	52.3	375	16	Q8XDR7	Q8XDR7 escherichia
8	45	52.3	2058	10	Q9LJG0	Q9LJG0 oryza sativ
9	44	51.2	317	5	Q96654	Q96654 trypanosoma
10	44	51.2	532	10	Q80941	Q80941 arabidopsis
11	44	51.2	689	5	Q76517	Q76517 trypanosoma
12	44	51.2	773	2	Q91U05	Q91U05 rhizobium m
13	43	50.0	312	5	Q17557	Q17557 caenorhabdi
14	43	50.0	398	10	Q94HS8	Q94HS8 oryza sativ
15	43	50.0	679	13	Q98856	Q98856 cynops pyrr
16	42	48.8	124	16	Q928F4	Q928F4 listeria in

17	42	48.8	184	11	Q9CWC3	Q9CWC3 mus musculu
18	42	48.8	316	5	O61129	O61129 ciliidia f
19	42	48.8	384	10	Q9LXJ7	Q9LXJ7 arabidopsis
20	42	48.8	416	10	Q9C6H1	Q9C6H1 arabidopsis
21	42	48.8	441	10	Q9FJ71	Q9FJ71 arabidopsis
22	42	48.8	451	10	Q9LXJ6	Q9LXJ6 arabidopsis
23	42	48.8	1065	2	Q9AHK8	Q9AHK8 borrelia bu
24	42	48.8	1417	5	Q9VSE1	Q9VSE1 drosophila
25	42	48.8	1858	16	Q9UIZ6	Q9UIZ6 agrobacteri
26	41.5	48.3	435	5	Q9TYG6	Q9TYG6 haemochus
27	41.5	48.3	435	5	Q9TZR3	Q9TZR3 haemochus
28	41.5	48.3	435	5	Q9GQW8	Q9GQW8 haemochus
29	41	47.7	260	16	P73030	P73030 synchocyst
30	41	47.7	271	10	Q22740	Q22740 arabidopsis
31	41	47.7	327	4	Q9NY49	Q9NY49 homo sapien
32	41	47.7	327	4	Q9NZP9	Q9NZP9 homo sapien
33	41	47.7	327	4	Q8WY68	Q8WY68 homo sapien
34	41	47.7	403	16	Q9RRZ6	Q9RRZ6 deinococcus
35	41	47.7	417	17	Q8TY81	Q8TY81 methanopyru
36	41	47.7	580	5	Q19873	Q19873 caenorhabdi
37	41	47.7	683	10	Q9LUR0	Q9LUR0 arabidopsis
38	41	47.7	738	5	Q9VRR0	Q9VRR0 drosophila
39	41	47.7	815	11	P70589	P70589 rattus norv
40	41	47.7	1187	11	P70475	P70475 rattus norv
41	40.5	47.1	96	12	Q9E873	Q9E873 gallid hep
42	40.5	47.1	96	12	Q9E865	Q9E865 gallid hep
43	40.5	47.1	295	8	O8SML8	O8SML8 carteria ol
44	40.5	47.1	658	16	P74702	P74702 synchocyst
45	40.5	47.1	1417	12	Q67631	Q67631 gallid hep

#### ALIGNMENTS

RESULT 1  
Q9Y6U5 PRELIMINARY; PRT; 264 AA.  
ID Q9Y6U5  
AC Q9Y6U5;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE WUGSC:H\_RG087B15.1 protein (fragment).  
GN WUGSC:H\_RG087B15.1  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OC NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99063792; PubMed=9847074;  
RA Sulston J.E., Waterston R.;  
RT "Toward a complete human genome sequence.";  
RL Genome Res. 8:1097-1108(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Strong C., Layman D., Graves T., Strommatt C.;  
RT "The sequence of Homo sapiens BAC clone CTB-87E15.";  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBD databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Waterston R.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBD databases.  
DR EMBL; Acc005061; Add43182.1; --  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 264 AA; 31203 MW; 35C9483003557B72 CRC64;  
FT NON\_TER 264  
FT NON\_TER 264  
Query Match 90.7%; Score 78; DB 4; Length 264;  
Best Local Similarity 86.7%; Pred. No. 1.7e-05;  
Matches 13; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 WKKKPRRLIEDDYL 15  
|||||

D	b	24	WKMKPRNLDDNDYL	38
RESULT 2				
ID	Q9GL50	PRELIMINARY;	PRT;	338 AA.
AC	Q9GL50;			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	Six transmembrane endothelial antigen of PAEC.			
OS	Sus scrofa (pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.			
OX	NCBI_TaxID=9823;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Nagaoka T., Boullay G., Coupel S., Coulon F., Teesson L.,			
RA	Hellan J.-M., Souillou J.-P., Charreau B.;			
RT	"Differential gene expression in endothelial cells during TNF-alpha-			
RT	and LPS-mediated activation."			
RL	Submitted (NCV-2000) to the EMBL/GenBank/DBJ databases.			
KW	EMBL; AFJ19659; AAC33868.1; -. Transmembrane.			
SO	SEQUENCE 338 AA; 39918 MW; ED490E86ED67A32B CRC64;			
Query Match		78.5%	Score 67.5;	DB 6; Length 338;
Best Local Similarity		86.7%;	Pred. No. 0.0015;	
Matches 13; Conservative 1; Mismatches 0; Indels 1; Gaps 14				
D	b	14	WKMKPRKRL-EDDYL	27
RESULT 3				
ID	Q9CWR7	PRELIMINARY;	PRT;	339 AA.
AC	Q9CWR7;			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	2410007B19Rik protein.			
CN	STAP OR 2410007B19RIK.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Aizawa K., Iwawaki M., Nishitani K., Kiyosawa H., Kondo S., Yamataka I.,			
RA	Saito T., Okazaki Y., Gojopori T., Bono H., Kasakawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Cavaletto T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,			
RA	Kuehl P., Lewis S., Matsumoto Y., Nakada I., Pesole G., Quackenbush J.,			
RA	Schirral L.M., Strubbi F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Botfield D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Butz C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.G., Kamuya M., Lee N.H.,			
RA	Lyon P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenfeld C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyko-Oka K., Wang K.H., Weitz C., Whiteaker C., Wilming L.,			
RA	Wysshew-Boris A., Yoshida K., Haegawa Y., Kawaji H., Kohntopp S.,			
RA	Hayashizaki Y.;			
RT	"Functional annotation of a full-length mouse cDNA collection."			
RL	Nature 409:685-690(2001). EMBL; AK010437; BAB26938.1; -.			

```

SO QY Db
SEQ MGI:1917608; Steap. 339 AA; 39264 MM; 3F7AB9C7520F0968 CRC64;
1 WKMKPRRLLEDDY 14
14 WKMKPRKGLLEDDSY 27

RESULT 4
ID AC 092422 PRELIMINARY; PRT; 339 AA.
092422
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Dudulin.
GN 1010001D01Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Serro V., Manivet P., Lamblin D., Vaubourdolle M., Kellermann O.,
RA Loric S.;
RT "Prostate and non-prostate expression of dudulin, the mouse ortholog
RT of human STEAP."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY029584; AAK50537.1; -.
DR MGI:1915678; 1010001D01Rik.
SQ SEQUENCE 339 AA; 39109 MM; 32A2C2F2B33BD0 CRC64;

Query Match
Best Local Similarity 73.3%; Score 63; DB 11; Length 339;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 WKMKPRRLLEDDY 14
14 WKMKPRKGLLEDDSY 27

Db 14 WKMKPRKGLLEDDSY 27

RESULT 5
ID AC 0924J9 PRELIMINARY; PRT; 339 AA.
0924J9
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Six-transmembrane epithelial antigen of the prostate.
GN STEAP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6;
RC MEDLINE=21371909; PubMed=11479226;
RA Yang D., Holt G.E., Velders M.P., Kwon E.D., Kast W.M.;
RT "Murine six-transmembrane epithelial antigen of the prostate, prostate
RT stem cell antigen, and prostate-specific membrane antigen: prostate-
RT transgenic cell-surface antigens highly expressed in prostate cancer of
RT transgenic adenocarcinoma mouse prostate mice."
RL Cancer Res. 61:5857-5860(2001).
DR EMBL; AF297098; AAK63126.1; -.
DR MGI:1917608; Steap.
RN Transmembrane.
SQ SEQUENCE 339 AA; 39318 MM; 4B26A71FF559E84F CRC64;

```

Query Match 73.3%; Score 63; DB 11; Length 339;  
 Best Local Similarity 71.4%; Pred. No. 0.009;  
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 WKMKPRRLNEEDDY 14  
 DB 14 WKMKPKGNLEDDSY 27

## RESULT 6

Q9UTAS PRELIMINARY; PRT; 237 AA.

AC Q9UTAS; 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Conserved hypothetical protein.  
 GN SPAC2588.15C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972H-;  
 RA Barrett B.G., Rajandream M.A., McDougall R.C., McLean J., Harris D.;  
 RL Submitted (NOV-1989) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL133225; CAB61781.1; -;  
 DR InterPro; IPR003827; DUF207.1.  
 DR Pfam; PF02676; DUF207.1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 237 AA; 26997 MW; 29ACFDD0C720997 CRC64;

Query Match 52.9%; Score 45.5; DB 3; Length 237;  
 Best Local Similarity 66.7%; Pred. No. 6.9;  
 Matches 10; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 2 KMKPRRLNEEDDY 15  
 DB 220 KMKPRRLNMDYL 234

## RESULT 7

Q8XDR7 PRELIMINARY; PRT; 375 AA.

AC Q8XDR7; 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE RNase D. processes tRNA precursor.  
 GN RND OR Z2847 OR EGS2513.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalaria E.T., Potamous K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
 RL Nature 409:529-533(2001).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,

RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shingawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 O157:H7 and genomic comparison with a laboratory strain K-12.";

RU DNA Res. 8:11-22(2001).  
 DR EMBL; AE005403; BA856793.1; -;  
 DR EMBL; AP002558; BAB35936.1; -;  
 DR InterPro; IPR002562; 3.5 exonuclease.  
 DR InterPro; IPR002121; HRDC.  
 DR Pfam; PF01612; 3.5 exonuclease; 1.  
 DR Pfam; PF00570; HRDC; 1.  
 DR SMART; SM00474; 35EXOC; 1.  
 DR SMART; SM00341; HRDC; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 375 AA; 42748 MW; 0D0F7D282602BAF CRC64;

Query Match 52.3%; Score 45; DB 16; Length 375;  
 Best Local Similarity 63.6%; Pred. No. 14;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 WKMKPRRLNEEDDY 11  
 DB 342 WKMKPNNLPE 352

## RESULT 8

Q9LGS0 PRELIMINARY; PRT; 2058 AA.

AC Q9LGS0; 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE P0406H10.14 protein.  
 GN P0406H10.14.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GAS) genomic DNA, chromosome 1, PAC  
 clone:P0406H10.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP002524; BAB07964.1; -;  
 SQ SEQUENCE 2058 AA; 226376 MW; 8A06B562B8BBD69 CRC64;

Query Match 52.3%; Score 45; DB 10; Length 2058;  
 Best Local Similarity 58.3%; Pred. No. 83;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 MKPRRLNEEDDY 14  
 DB 496 LKGRKNISEDDY 507

## RESULT 9

Q96654 PRELIMINARY; PRT; 317 AA.

AC Q96654; 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-MAY-2002 (TREMBlrel. 20, Last annotation update)  
 DE Activated protein kinase C receptor homolog.  
 OS Trypanosoma vivax.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5699;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IIDAT1.2;

RA Djikeng A., Majiwa P.A.;  
 RT "Trypanosoma vivax homologue of the receptor for activated protein  
 kinase C."  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 7 MD REPEATS (TRP-ASP DOMAINS).  
 DR EMBL; AF100287; AAC72850.1; -.  
 DR InterPro; IPR01680; WD40.  
 DR Pfam; PF00400; WD40; 7.  
 DR PRINTS; PR00320; GPROTEINRPT.  
 DR ProDom; PD000016; WD40; 4.  
 DR SMART; SM00320; WD40; 7.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_3.  
 DR PROSITE; PS50082; WD\_REPEATS\_2; 5.  
 DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
 KW Kinase; Receptor; Repeat; WD repeat.  
 SQ SEQUENCE 317 AA; 34895 MW; B3CFB35D520AB31C CRC64;

Query Match 51.2%; Score 44; DB 5; Length 317;  
 Best Local Similarity 50.0%; Pred. No. 17;  
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 WKMKPRRLLEDDY 14  
 DB 43 WSDNPNRHAENDY 56

## RESULT 10

ID 080941 PRELIMINARY; PRT; 532 AA.  
 AC 080941;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE A12377730 protein.  
 GN AT2377730.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC euroside II; Brassicales; Brassicaceae; Arabidopsids.  
 NCBI\_TaxID=3702;

RA [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shae T.P., Benito M.-I., Town C.D.,  
 Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
 Beil C.R., Kechum K.A., Lee U.-J., Roming C.M., Koo H., Moffat K.S.,  
 Cronin L.A., Shen M., Vanaken S.B., Umayam L., Taiton L.J., Gill J.E.,  
 Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 Copenhaver G.P., Preuss D., Nieman W.C., White O., Eisen J.A.,  
 Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 thaliana."  
 RL Nature 402:761-768(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Lin X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC004684; AAC23643.1; -.  
 SQ SEQUENCE 532 AA; 60372 MW; 15706686846B4B5 CRC64;

Query Match 51.2%; Score 44; DB 10; Length 532;  
 Best Local Similarity 55.6%; Pred. No. 29;  
 Matches 10; Conservative 0; Mismatches 2; Indels 6; Gaps 1;

QY 1 WKMKPRRLLEDDY 12  
 DB 476 WKAPRRQCEIVNSFD 493

RESULT 11  
 076517

ID 076517 PRELIMINARY; PRT; 659 AA.  
 AC 076517;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Lysosomal/endosomal membrane protein p67.  
 OS Trypanosoma brucei rhodesiense.  
 OC Eukaryota; Eukaryozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 NCBI\_TaxID=31286;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99151764; PubMed=10029306;  
 RA Kelley R.J., Alexander D.L., Cowan C., Balber A.E., Bangs J.D.;  
 RT "Molecular cloning of p67, a lysosomal membrane glycoprotein from  
 Trypanosoma brucei."  
 RL Mol. Biochem. Parasitol. 98:17-28(1999).  
 DR EMBL; AF074867; AAC28456.1; -.  
 DR InterPro; IPR01064; Crystallin.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; UNKNOWN\_1.  
 SQ SEQUENCE 659 AA; 72704 MW; 4E96EE5C81C06993 CRC64;

Query Match 51.2%; Score 44; DB 5; Length 659;  
 Best Local Similarity 61.5%; Pred. No. 37;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 MKPRRLLEDDYL 15  
 DB 635 MRPRRTEDELL 647

RESULT 12  
 ID 091005 PRELIMINARY; PRT; 773 AA.  
 AC 091005;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE TOP protein.  
 GN TOP.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 NCBI\_TaxID=382;  
 RA [1]  
 RP SEQUENCE FROM N.A.  
 RC TRANSPOSON-MERCURY RESISTANCE TRANSPOSON TN5718;  
 RA Schneiker S., Keller M., Droege M., Lanka E., Puhler A.,  
 Selbitschka W.;  
 RT "The genetic organization and evolution of the broad-host-range  
 mercury resistance plasmid PSB102 isolated from a microbial population  
 residing in the rhizosphere of alfalfa."  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ034453; CAC79182.1; -.  
 DR InterPro; IPR002936; DNAPrim\_toprim.  
 DR InterPro; IPR000380; Prok\_cpisomraae.  
 DR Pfam; PF01131; Topoisom\_bac; 1.  
 DR Pfam; PF01751; Toprim; 1.  
 DR TIGRfam; TIGR01051; topa\_bact; 1.  
 KW Plasmid.  
 SQ SEQUENCE 773 AA; 86188 MW; DBE36F5AC6F20D68 CRC64;

Query Match 51.2%; Score 44; DB 2; Length 773;  
 Best Local Similarity 53.3%; Pred. No. 44;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 WKMKPRRLLEDDYL 15  
 DB 99 WHLKESLIGEDDYL 113

RESULT 13  
 017557



FT	SIGNAL	1	19	BY SIMILARITY.
FT	PROPEP	20	107	ACTIVATION PEPTIDE (BY SIMILARITY).
FT	CHAIN	108	679	92 KDA TYPE IV COLLAGENASE.
FT	DOMAIN	98	105	AUTOINHIBITOR REGION.
FT	DOMAIN	224	281	FIBRONECTIN TYPE-II.
FT	DOMAIN	282	339	FIBRONECTIN TYPE-II.
FT	DOMAIN	340	397	FIBRONECTIN TYPE-II.
FT	DOMAIN	486	679	HEMOPEXIN-LIKE.
FT	METAL	401	401	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT SITE	402	402	BY SIMILARITY.
FT	METAL	405	405	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	411	411	ZINC (CATALYTIC) (BY SIMILARITY).
FT	DISULFID	491	676	BY SIMILARITY.
FT	CARBOHYD	82	82	N-LINKED (GLCNAC..) (POTENTIAL).
SO	SEQUENCE	679 AA;	75581 MW;	573CDD9DBP85524 CRC64;

Query Match 50.0%; Score 43; DB 13; Length 679;  
 Best Local Similarity 40.0%; Pred. No. 57;  
 Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 WKMKPRRLLEDDYL 15  
 DB 655 WKMTPRKQVDQGVV 669

Search completed: March 26, 2003, 16:49:42  
 Job time : 20.6977 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 16:38:14 ; Search time 4.36047 Seconds  
(without alignments)  
142.679 Million cell updates/sec

Title: US-10-010-667A-2\_COPY\_14\_28  
Perfect score: 86  
Sequence: 1 WKMKRRNIEBDYL 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred: No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	86	100.0	339 1 STEA_HUMAN	Q9uh68 homo sapien
2	45	52.3	375 1 RND_ECOLI	P09155 escherichia
3	44	51.2	193 1 VANI_VIBAN	P74945 vibrio angu
4	43	50.0	539 1 PRRI_SCHPO	O14283 schizosacch
5	42	48.8	441 1 ODP2_BACSU	P21883 bacillus su
6	42	48.8	847 1 AGUA_TREIRE	Q99024 trichoderma
7	42	48.8	1018 1 ST31_MOUSE	Q99m11 mus musculu
8	42	48.8	1019 1 ST31_HUMAN	Q93x11 homo sapien
9	40.5	47.1	639 1 HTD3_TREPA	O83949 treponema p
10	40.5	47.1	1415 1 ICP4_HBVMG	Q02362 marck's dis
11	40	46.5	36 1 RET4_CHICK	P30370 gallus gall
12	40	46.5	137 1 RET3_XENLA	P50568 xenopus lae
13	40	46.5	197 1 I1B8_HUMAN	O95998 homo sapien
14	40	46.5	199 1 SYR_ARCFU	O29368 archaeglob
15	40	46.5	553 1 GPT1_CANAL	O74248 candida alb
16	39.5	45.9	607 1 HTD3_FUSN1	O8rgh4 fusobacteri
17	39	45.3	166 1 VSN1_NOCAB	P50186 nocardia ae
18	39	45.3	347 1 IDH1_MERJA	Q58891 metanococc
19	39	45.3	457 1 IRF7_MOUSE	P70434 mus musculu
20	39	45.3	20 YU20_DROME	Q9vif0 drosophila
21	39	45.3	1469 1 DP27_CAEEL	P48996 caenorhabdi
22	38	44.2	193 1 LUXJ_VIBFI	P12747 vibrio fisc
23	38	44.2	345 1 LEU3_LACIA	O02143 lactococcus
24	38	44.2	352 1 YG35_MYCPN	P75162 mycoplasma
25	38	44.2	370 1 HAO1_HUMAN	Q9vum8 homo sapien
26	38	44.2	427 1 ODP2_BACST	P18611 bacillus st
27	38	44.2	691 1 VP80_NPVAC	O00733 autographa
28	38	44.2	725 1 YGL4_YEAST	P53134 saccharomyc
29	38	44.2	729 1 KEX1_YEAST	P09620 saccharomyc
30	38	44.2	736 1 YBL7_YEAST	P38623 saccharomyc
31	38	44.2	839 1 YVDB_HAEIN	P45182 haemophilus
32	38	44.2	925 1 RNT1_SCHPO	Q09880 schizosacch
33	38	44.2	968 1 PKD2_HUMAN	O13563 homo sapien

34	38	44.2	1371 1 YOL4_CAEEL	Q02331 caenorhabdi
35	38	44.2	3866 1 HRX_MOUSE	P55200 mus musculu
36	38	44.2	3969 1 HRX_HUMAN	Q03164 homo sapien
37	37.5	43.6	716 1 ENPL_RABIT	O18750 oryctolagus
38	37.5	43.6	795 1 ENPL_CHICK	P08110 gallus gall
39	37.5	43.6	802 1 ENPL_MOUSE	P08113 mus musculu
40	37.5	43.6	803 1 ENPL_HUMAN	P14623 homo sapien
41	37	43.0	133 1 HEMI_PYRAP	P37363 pyrrhocoris
42	37	43.0	136 1 RET3_FUGRU	O42386 fuigu rubrip
43	37	43.0	170 1 GWA2_STRCO	Q9rjb4 streptococc
44	37	43.0	193 1 LUXJ_VIBFI	P35328 vibrio fisc
45	37	43.0	230 1 VIF_TILVI	P23430 viena lenti

## ALIGNMENTS

RESULT 1  
ID STEA\_HUMAN STANDARD; PRT; 339 AA.  
AC Q9UH68; 095034;  
DT 16-OCT-2001 (Rel. 40, last sequence update)  
DT 16-OCT-2001 (Rel. 40, last sequence update)  
DT 15-JUN-2002 (Rel. 41, last annotation update)  
DE Six transmembrane epithelial antigen of prostate.  
GN STEAP OR STEAP1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20056277; PubMed=10588738;  
RA Hubert R.S., Vivanco I., Chen E., Raastegar S., Leong K.,  
RA Mitchell S.C., Madraswala R., Zhou Y., Kuo J., Raitano A.B.,  
RA Jakobovits A., Salfran D.C., Afari D.E.H.;  
RT "STEAP: A Prostate-Specific Cell-Surface Antigen Highly Expressed in  
RT human prostate tumors.";  
RT Proc. Natl. Acad. Sci. U.S.A. 96:14523-14528 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Abu-Threideh J., Stoneking T., Langston Y., Maupin R.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=skin;  
RA Strauberg R.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PROSTATE TUMORS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; AF186249; AAF17479.1; -;  
DR EMBL; AC005053; AAC79150.1; ALT INIT.  
DR EMBL; AC004969; AAD15620.1; ALT INIT.  
DR EMBL; BC011802; AAH11802.1; -;  
DR Genew; HGNC:11378; STEAP.  
DR MIM; 604415; -;  
KW Transmembrane; Antigen.  
FT TRANSMEM 71 91  
FT TRANSMEM 119 139  
FT TRANSMEM 164 184  
FT TRANSMEM 218 238  
FT TRANSMEM 258 278  
FT TRANSMEM 291 311  
SQ SEQUENCE 339 AA; 39851 MW; 55443A170C870387 CRC64;

Query Match 100.0%; Score 86; DB 1; Length 339;  
 Best Local Similarity 100.0%; Pred. No. 3; le-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WKMKPRNLEEDYL 15  
 Db 14 WKMKPRNLEEDYL 28

RESULT 2  
 RND\_ECOLI STANDARD; PRT; 375 AA.  
 AC P09155;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ribonuclease D (EC 3.1.26.3) (RNase D).  
 GN RND OR B1804.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 NCBI\_TaxID=562;  
 (1)  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-6.  
 RX MEDLINE=8289400; PubMed=3041371;  
 RA Zhang J., Deutscher M.P.;  
 RT "Escherichia coli RNase D: sequencing of the rnd structural gene and  
 purification of the overexpressed protein."  
 RL Nucleic Acids Res. 16:6265-6278(1988).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=92278503;  
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Berna N.T., Burland V.,  
 RA Riley M., Colisdo-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12."  
 RL Science 277:1233-1238(1997).  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97251358; PubMed=9097040;  
 RA Itoh K., Kawai H., Baba T., Fujita K., Hayashi K., Inada T.,  
 RA Isono K., Kawai H., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Nakano K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,  
 RA Saito N., Sempel G., Seki Y., Sivasubramanian S., Tagami H.,  
 RA Takeda Y., Takemoto K., Wada C., Yamamoto Y., Horikuchi T.;  
 RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 40.1-50.0 min region on the linkage map."  
 RL DNA Res. 3:379-392(1996).  
 RN (4)  
 RP SEQUENCE OF 1-38 FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=94150456; PubMed=8107670;  
 RA Fulda M., Heinz E., Wolter F.P.;  
 RT "The fadD gene of Escherichia coli K12 is located close to rnd at  
 39.6 min of the chromosomal map and is a new member of the  
 AMP-binding protein family."  
 RL Mol. Genet. 242:241-249(1994).  
 CC -1- FUNCTION: CLEAVES MULTIMERIC TRNA PRECURSOR AT THE SPACER REGION.  
 CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-  
 phosphononucleoside.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial/  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC EMBL: X07055; CAA30098.1; -  
 CC DR EMBL: AE000274; AAC74874.1; -  
 CC DR EMBL: D90823; BAA15599.1; -  
 CC DR EMBL: D90824; BAA15608.1; -  
 CC DR EMBL: D90825; BAA15613.1; -  
 CC DR EMBL: X70994; CAA50322.1; -  
 CC DR PIR: S01223; NRECD.  
 CC DR EcoGene: EG10858; rnd.  
 CC DR InterPro: IPR002562; 3\_5\_exonuclease.  
 CC DR InterPro: IPR002121; HRDC.  
 CC DR Pfam: PF00570; HRDC; 1.  
 CC DR Pfam: PF01612; 3\_5\_exonuclease; 1.  
 CC DR SMART: SM00474; 35EXOC; 1.  
 CC DR SMART: SM00341; HRDC; 1.  
 CC KM HydroLase; Nuclease; Exonuclease; tRNA processing; Complete proteome.  
 CC SQ SEQUENCE 375 AA; 42734 MW; AC0B6D883712BAE CRC64;

Query Match 52.3%; Score 45; DB 1; Length 375;  
 Best Local Similarity 63.6%; Pred. No. 3.2;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 WKMKPRNLEE 11  
 Db 342 WKMKPRNLEE 352

RESULT 3  
 VANI\_VIBAN STANDARD; PRT; 193 AA.  
 ID VANI\_VIBAN  
 AC P74945;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Autoinducer synthesis protein vanti.  
 GN VANI.  
 OS Vibrio anguillarum (Listonella anguillarum).  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Listonella.  
 OX NCBI\_TaxID=55601;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NB10 / Serotype O1;  
 RX MEDLINE=97284506; PubMed=9139920;  
 RA Milton D.L., Hardman A., Camara M., Chhabra S.R., Bycroft B.W.,  
 RA Stewart G.S.A.B., Williams P.;  
 RT "Quorum sensing in Vibrio anguillarum: characterization of the  
 RT vanti/vanr locus and identification of the autoinducer N-(3-  
 RT oxodecanoyl)-L-homoserine lactone."  
 RL J. Bacteriol. 179:3004-3012(1997).  
 CC -1- FUNCTION: REQUIRED FOR THE SYNTHESIS OF N-(3-OXODECANOYL)-L-  
 CC HOMOSERINE LACTONE (ODHL), AN AUTOINDUCER MOLECULE WHICH BINDS TO  
 CC VANI.  
 CC -1- SIMILARITY: BELONGS TO THE AUTOINDUCER SYNTHETASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial/  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC EMBL: U69677; AAC45212.1; -  
 CC DR InterPro: IPR001690; Autoind\_synth.  
 CC DR Pfam: PF00765; Autoind\_synth; 1.  
 CC DR PRINTS: PR01549; AUTOINDCRSYN.  
 CC DR PRODOM: PD002752; Autoind\_synth; 1.  
 CC DR PROSITE: PS00949; AUTOINDUCERS\_SYNTH; 1.  
 CC KM Quorum sensing; Autoinducer synthesis.  
 CC SQ SEQUENCE 193 AA; 22127 MW; 238E760CD8701188 CRC64;

Query Match 51.2%; Score 44; DB 1; Length 193;

```

Oy      1 WAKPRRNLEEDDY 14
        | : | | : |
Db      35 WELKTRNGMETDEY 48

RESULT 4
PRRI_SCHPO STANDARD; PRT; 539 AA.
ID PRRI SCHPO
AC 014283; OSUTX5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription factor prri (Pombe response regulator 1).
DE PRRI OR SPAC8C9.14.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RX MEDLINE=99279565; PubMed=10348908;
RT Omhaya R., Kato C., Yamada H., Aiba H., Mizuno T.;
RT A fission yeast gene (prri(+)) that encodes a response regulator
RT implicated in oxidative stress response.";
RL J. Biochem. 125:1061-1066(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornbly T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leacher S., McDonald S., McLean U.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymoult B.,
RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreno S., Gloux S., Lelaire V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garçon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,
RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potlatchkin J.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;
RL "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
RN [3]
RP SEQUENCE OF 224-413 FROM N.A.
RX STRAIN=968 h90;
RX MEDLINE=20223868; PubMed=10759689;
RA Ding D.Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
RA Hiroka Y.;
RT "Large-scale screening of intracellular protein localization in living
RT fission yeast cells by the use of a GFP-fusion genomic DNA library.";
RL Genes Cells 5:169-190(2000).
CC -1- FUNCTION: INVOLVED IN OXIDATIVE STRESS. TRANSCRIPTION FACTOR THAT
CC ACTS UPON TRR1 AND CTT1.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE HSF FAMILY.
CC -1- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk.)
CC -----
CC
CC EMBL; AB041768; BAB16722.1; -
CC
CC EMBL; Z99168; CAB16301.1; -
CC
CC EMBL; AB027943; BAA87247.1; -
CC
CC HSSP; P22121; 2HTS.
CC
CC InterPro; IPR000232; HSF_DNA_bind.
CC
CC InterPro; IPR002341; HSF_ETS.
CC
CC InterPro; IPR001789; Response_reg.
CC
CC Pfam; PF00072; response_reg; 1.
CC
CC Pfam; PF00447; HSF_DNA_bind; 1.
CC
CC PRINTS; P00056; HSFDOMAIN.
CC
CC ProDom; PD000039; Response_reg; 1.
CC
CC ProDom; PD001788; HSF_DNA_bind; 1.
CC
CC SMART; SM00415; HSF; 1.
CC
CC SMART; SM00448; REC; 1.
CC
CC PROSITE; PS00434; HSF_DOMAIN; 1.
CC
CC PROSITE; PS50110; RESPONSE_REGULATORY; 1.
CC
CC Transcription regulation; Sensory transduction; Nuclear protein;
CC DNA-binding; Phosphorylation.
CC
CC FT DOMAIN 7 111 RESPONSE_REGULATORY.
CC FT MOD_RES 369 483 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 418 418 PHOSPHORYLATION (BY SIMILARITY).
CC
CC SEQ SEQUENCE 539 AA; 60045 MW; F3B4945C59582871 CRC64;
CC
CC
CC Query Match 50.0%; Score 43; DB 1; Length 539;
CC Best Local Similarity 60.0%; Pred. No. 10;
CC Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
CC
CC
CC QY 1 WKMKPRNLIEDYLV 15
CC
CC DB 364 WKOPRILVDEDEL 378
CC
CC
CC RESULT 5
CC CDP2_BACSU
CC ID_ODP2_BACSU STANDARD; PRT; 441 AA.
CC
CC AC P21883;
CC DT 01-MAY-1991 (Rel. 18, Created)
CC DT 01-MAY-1991 (Rel. 18, Last sequence update)
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC DE Dihydroisopantide acetyltransferase component of pyruvate dehydrogenase
CC complex (EC 2.3.1.12) (E2) (S complex, 48 kDa subunit).
CC DN PDCB GN CECB.
CC OS Bacillus subtilis.
CC OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
CC OX NCBI_TaxId=1423;
CC
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC
CC RC STRAIN=168;
CC RX MEDLINE=90368558; PubMed=1697575;
CC RA Hemila H., Palva A., Paulin L., Arvidson S., Palva I.;
CC RT "Seretary S complex of Bacillus subtilis: sequence analysis and
CC identity to pyruvate dehydrogenase."
CC RL J. Bacteriol. 172:5052-5063(1990).
CC
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC
CC RC STRAIN=168;
CC RX MEDLINE=97124187; PubMed=8969500;
CC RA Winters P., Caldwell R., Enfield L., Ferrari E.;
CC RT "The ampS-nprE (124 degrees-127 degrees) region of the Bacillus
CC subtilis 168 chromosome: sequencing of a 27 kb segment and
CC identification of several genes in the area."
CC RL Microbiology 142:3033-3037(1996).
CC
CC RP SEQUENCE FROM N.A.

```

CC STRAIN=168;  
 RA Caldwell R.M., Ferrari E.;  
 RT "Sequence analysis of the mobA-empS region of the *Bacillus subtilis*  
 chromosome";  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 CC  
 CC -1- FUNCTION: The pyruvate dehydrogenase complex catalyzes the overall  
 conversion of pyruvate to acetyl-CoA and CO(2). It contains  
 multiple copies of three enzymatic components: pyruvate  
 dehydrogenase (E1), dihydrolipoamide acetyltransferase (E2) and  
 lipoamide dehydrogenase (E3).  
 CC  
 CC -1- FUNCTION: THE B SUBUNIT IS PDH COMPLEX POSSESSORS ALSO BRANCHED-CHAIN  
 2-OXOACID DEHYDROGENASE (BCDH) ACTIVITY.  
 CC  
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + dihydrolipoamide = CoA + S-  
 acetyldihydrolipoamide.  
 CC  
 CC -1- COFACTOR: THE E2 COMPONENT CONTAINS ONE COVALENTLY-BOUND LIPOYL  
 COFACTOR.  
 CC  
 CC -1- SUBUNIT: FORMS A 24-POLYPEPTIDE STRUCTURAL CORE WITH OCTAHEDRAL  
 SYMMETRY.  
 CC  
 CC -1- SIMILARITY: BELONGS TO THE 2-OXOACID DEHYDROGENASE FAMILY.  
 CC  
 CC -1- SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL: MS7435; AAA62683.1; -;  
 CC EMBL: AF012285; AAC24934.1; -;  
 CC EMBL: Z99111; CAB13333.1; -;  
 CC PIR: D36718; D36718.  
 CC HSSP: P11961; 2PDS.  
 CC Subtilist; BG10209; pdhc.  
 CC InterPro: IPR001078; 2Oxocacid dh.  
 CC InterPro: IPR000089; Biotin liPOL.  
 CC InterPro: IPR004167; E3 binding.  
 CC InterPro: IPR003016; LiPOL.  
 CC Pfam: PF00198; 2-oxocacid dh; 1.  
 CC Pfam: PF02817; e3\_binding; 1.  
 CC Pfam: PF00164; biotin liPOL; 1.  
 CC ProDom: PD00115; 2Oxocacid dh; 1.  
 CC PROSITE: PS00189; LiPOL; 1.  
 CC KMW Glycolysis; Transferrase; Acyltransferase; LiPOL; Complete proteome.  
 CC INIT MET 0  
 CC DOMAIN 1 85 LIPOYL BINDING.  
 CC FT DOMAIN 139 181 E1/E3 BINDING.  
 CC FT BINDING 42 42 LIPOYL.  
 CC FT ACT SITE 412 412 POTENTIAL.  
 CC SQ SEQUENCE 441 AA; 47407 MW; 6B44D47879D134DF CRC64;  
 Query Match 48.8%; Score 42; DB 1; Length 441;  
 Best Local Similarity 46.7%; Pred. No. 12;  
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 WKMKPRRNLEDDYL 15  
 DB 21 WPKVKNDEVDV 35  
 RESULT 6  
 AGUA TRIRE STANDARD; PRT; 847 AA.  
 AC 099074;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Alpha-glucuronidase precursor (EC 3.2.1.139) (Alpha-glucosiduronase)  
 DE (GIRI)  
 OS Trichoderma reesei (Hypocrea jecorina).  
 CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 CC Hypocreales; Hypocreaceae; Hypocrea;

CC NCB1\_TaxID=51453;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=OM9414 / Rut C-30;  
 RX MEDLINE=96527277; PubMed=8654984;  
 RA Margolis-Clark E., Saloheimo M., Silka-Aho M., Penttilae M.;  
 RT "The alpha-glucuronidase-encoding gene of *Trichoderma reesei*";  
 RL Gene 172:171-172(1996).  
 CC  
 CC -1- FUNCTION: RELEASES 4-O-METHYLGLUCURONIC ACID FROM XYLAN.  
 CC  
 CC -1- CATALYTIC ACTIVITY: An alpha-D-glucuronoside + H(2)O = an alcohol  
 + D-glucuronate.  
 CC  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC  
 CC -1- SIMILARITY: BELONGS TO FAMILY 67 OF GLYCOSYL HYDROLASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL: Z68706; CAA92949.1; -;  
 CC InterPro: IPR005154; Glyco\_hydro\_67.  
 CC Pfam: PF03648; Glyco\_hydro\_67; 1.  
 CC KMW Xylan degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.  
 CC SIGNAL 1 19  
 CC FT CHAIN 20 847  
 CC FT CARBOHYD 52 52 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC FT CARBOHYD 238 238 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC FT CARBOHYD 321 321 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC FT CARBOHYD 353 353 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC FT CARBOHYD 586 586 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC FT CARBOHYD 692 692 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC FT CARBOHYD 740 740 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC FT CARBOHYD 767 767 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC SQ SEQUENCE 847 AA; 93424 MW; 214AAE8782FDB4 CRC64;  
 Query Match 48.8%; Score 42; DB 1; Length 847;  
 Best Local Similarity 53.8%; Pred. No. 26;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 WKMKPRRNLEDDYL 14  
 DB 108 KLSPKLNKEDGY 120  
 RESULT 7  
 ST31\_MOUSE STANDARD; PRT; 1018 AA.  
 AC 099M11;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Serine/threonine protein kinase 31 (EC 2.7.1.37).  
 GN STK31.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CC NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=21175748; PubMed=11279525;  
 RA Wang P.J., McCarrey J.R., Yang F., Page D.C.;  
 RT "An abundance of X-linked genes expressed in spermatogonia";  
 RL Nat. Genet. 27:422-426(2001).  
 CC  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
 CC  
 CC -1- TISSUE SPECIFICITY: Testis-specific. Expressed only in male germ  
 cells.  
 CC  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC  
 CC -1- SIMILARITY: CONTAINS 1 TUDOR DOMAIN.  
 CC -----

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AF285580; AKK1959.1; -  
DR MGD; MG1:1924735; STK31.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR001097; Maternal\_tudor.  
DR InterPro; IPR000080; S\_nase.  
DR InterPro; IPR002290; Ser\_thr\_pkinase.  
DR InterPro; IPR002999; Tudor.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF00567; TUDOR; 1.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00220; S\_TKC; 1.  
DR SMART; SM00333; TUDOR; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS50304; TUDOR; 1.  
DR Transferrase; Serine/threonine-protein kinase; ATP-binding;  
KW Colled coil.  
KW DOMAIN 78 137 TUDOR.  
FT DOMAIN 105 245 TNASE-LIKE.  
FT DOMAIN 298 358 COILED COIL (POTENTIAL).  
FT DOMAIN 711 1018 PROTEIN\_KINASE.  
FT NP\_BIND 717 725 ATP (BY SIMILARITY).  
FT BINDING 738 738 ATP (BY SIMILARITY).  
SQ SEQUENCE 1018 AA; 115027 MW; 1206284ED4E90816 CRC64;  
  
Query Match 48.8%; Score 42; DB 1; Length 1018;  
Best local Similarity 53.8%; Pred. No. 31;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 WKMKPRRLNEEDD 13  
Db 646 WKLVKSNLEESD 658  
||:|||||  
-----  
RESULT 8  
ST31\_HUMAN STANDARD; PRT; 1019 AA.  
AC Q9BXU1; Q9BXU8; 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Serine/threonine protein kinase 31 (EC 2.7.1.37) (Serine/threonine-  
DE protein kinase NTP-SPK).  
GN STK31.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=21175748; PubMed=11279525;  
RA Wang P.J., McCarrey J.R., Yang F., Page D.C.;  
RT "An abundance of X-linked genes expressed in spermatogonia.";  
RL Nat. Genet. 27:422-426(2001).  
[2]  
RP SEQUENCE OF 24-1019 FROM N.A.  
RC TISSUE=Testis;  
RA Zhou Z.M.;  
RT "Cloning of a new protein kinase gene related to human testis  
RT development.";  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ database.  
CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
CC -1- TISSUE SPECIFICITY: testis-specific.  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC -1- SIMILARITY: CONTAINS 1 TUDOR DOMAIN.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AF285599; AKK1978.1; -  
DR EMBL; AF332194; AKK17193.1; -  
DR Genew; HGNC:11407; STK31.  
DR MIM; 605790; -  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR001097; Maternal\_tudor.  
DR InterPro; IPR000080; S\_nase.  
DR InterPro; IPR002290; Ser\_thr\_pkinase.  
DR InterPro; IPR002999; Tudor.  
DR Pfam; PF00069; pkinase; 2.  
DR Pfam; PF00567; TUDOR; 2.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00220; S\_TKC; 1.  
DR SMART; SM00333; TUDOR; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; FALSE\_NEG.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS50304; TUDOR; 1.  
KW Transferrase; Serine/threonine-protein kinase; ATP-binding;  
KW Colled coil.  
KW DOMAIN 78 137 TUDOR.  
FT DOMAIN 105 242 TNASE-LIKE.  
FT DOMAIN 298 355 COILED COIL (POTENTIAL).  
FT DOMAIN 710 1019 PROTEIN\_KINASE.  
FT NP\_BIND 716 724 ATP (BY SIMILARITY).  
FT BINDING 737 737 ATP (BY SIMILARITY).  
FT CONFICT 300 300 I -> F (IN REF. 2).  
FT CONFICT 509 509 F -> Y (IN REF. 2).  
FT CONFICT 621 623 KRI -> NKS (IN REF. 2).  
FT CONFICT 715 715 Y -> C (IN REF. 2).  
FT CONFICT 820 820 V -> A (IN REF. 2).  
FT CONFICT 948 948 K -> I (IN REF. 2).  
FT CONFICT 963 963 A -> G (IN REF. 2).  
FT CONFICT 1000 1000 T -> P (IN REF. 2).  
FT CONFICT 1010 1010 P -> T (IN REF. 2).  
SQ SEQUENCE 1019 AA; 115729 MW; 5B58244A980D5590A CRC64;  
  
Query Match 48.8%; Score 42; DB 1; Length 1019;  
Best local Similarity 53.8%; Pred. No. 31;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 WKMKPRRLNEEDD 13  
Db 645 WKLVKSNLEESD 657  
||:|||||  
-----  
RESULT 9  
HTPG\_TREPA STANDARD; PRT; 639 AA.  
AC HTPG\_TREPA  
ID O83949;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Chaperone protein htpg (Heat shock protein htpg) (high temperature  
DE protein G).  
GN HTPG OR TP0984.  
OS Treponema pallidum.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
OX NCBI\_TaxID=160;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nichols;  
RX MEDLINE=98332770; PubMed=9665876;  
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,

```

RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salberg S., Peterson J.,
RA Khakl H., Richardson D., Howell J.K., Chidambaram M., Uterback T.,
RA McDonald L., Attlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman U.J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete."
RL Science 281:375-388(1998).
CC -1- FUNCTION: Molecular chaperone. Has ATPase activity (By
CC similarity).
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE001265; AAC65938.1; -.
CC HSSP: P02829; IAH8.
CC TIGR: TP0984; -.
CC InterPro: IPR003594; ATPbind_ATPase.
CC InterPro: IPR001404; Hsp90.
CC Pfam: PF00183; HSP90; 2.
CC Pfam: PF02518; HATPase_C; 1.
CC PRINTS: PR00775; HEATSHOCK90.
CC SMART: SM00387; HATPase_C; 1.
CC PROSITE: PS00298; HSP90; 1.
CC Chaperone; ATP-binding; Heat shock; Complete proteome.
CC DOMAIN 1 348 A; SUBSTRATE-BINDING (By similarity).
CC FT DOMAIN 349 565 B (By similarity).
CC FT DOMAIN 566 639 C.
CC SQ SEQUENCE 639 AA; 72937 MW; 3E8FDBAC2282C31D CRC64;

Query Match 47.1%; Score 40.5; DB 1; Length 639;
Best Local Similarity 50.0%; Pred. No. 33;
Matches 7; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

QY 1 WKMKPRRLNLEDDY 14
Db 246 WK-RPKSELKEDY 258

RESULT 10
ICP4 HSVMG STANDARD; PRT; 1415 AA.
AC 002362;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Trans-acting transcriptional activator protein ICP4 (Immediate-early
DE protein IE175).
GN ICP4.
OS Marek's disease herpesvirus (strain GA) (MDHV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Marek's disease-like viruses.
OX NCB1_TaxID=10388; [1]
RN NCB1_TaxID=10388; [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92351564; PubMed=1322594;
RA Anderson A.S., Francesconi A., Morgan R.W.;
RT "Complete nucleotide sequence of the Marek's disease virus ICP4
RT gene."
RL Virology 189:667-667(1992).
CC -1- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS. IT IS
CC REQUIRED FOR THE SWITCH FROM IMMEDIATE-EARLY TO EARLY MODE OF

```

```

CC GENE EXPRESSION.
CC -1- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC -1- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF
CC PHOSPHORYLATION.
CC -1- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M75729; AAA46111.1; -.
CC PIR: A42991; EDBEGA.
CC InterPro: IPR005205; Herpes_ICP4_C.
CC InterPro: IPR005206; Herpes_ICP4_N.
CC Pfam: PF03584; Herpes_ICP4_N; 1.
CC Pfam: PF03585; Herpes_ICP4_C; 1.
CC KM Early protein; Transcription regulation; Trans-acting factor;
CC DNA-binding; Phosphorylation; Nuclear protein.
CC FT DOMAIN 155 200
CC SQ SEQUENCE 1415 AA; 154936 MW; C0846F7BEFAD1126 CRC64;

Query Match 47.1%; Score 40.5; DB 1; Length 1415;
Best Local Similarity 60.0%; Pred. No. 81;
Matches 9; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 2 KMKPR-RNLEDDYL 15
Db 222 KASPRRKLEDDYL 236

RESULT 11
RET4 CHICK STANDARD; PRT; 36 AA.
AC RET4_CHICK
AC P30370;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Retinoid acid-binding protein II, cellular (CRABP-II) (Fragment).
GN CRABP2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCB1_TaxID=9031; [1]
RN NCB1_TaxID=9031; [1]
RP SEQUENCE.
RX MEDLINE=89087453; PubMed=2849937;
RA Kitamoto T., Momoi T., Momoi M.;
RT "The presence of a novel cellular retinoic acid-binding protein in
RT chick embryos: purification and partial characterization."
RL Biochem. Biophys. Res. Commun. 157:1302-1308(1988).
GN [2]
RP ERRATUM.
RA Kitamoto T., Momoi T., Momoi M.;
RL Biochem. Biophys. Res. Commun. 159:371-371(1989).
CC -1- FUNCTION: CYTOSOLIC CRABPS MAY REGULATE THE ACCESS OF RETINOIC
CC ACID TO THE NUCLEAR RETINOIC ACID RECEPTORS.
CC -1- FUNCTION: CRABP-II MAY PARTICIPATE IN A REGULATORY FEEDBACK
CC MECHANISM TO CONTROL THE ACTION OF RETINOIC ACID ON CELL
CC DIFFERENTIATION.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: EMBRYO.
CC -1- SIMILARITY: BELONGS TO THE FABP/P2/CRBP/CRABP FAMILY OF
CC TRANSPORTERS.
CC PIR: B31872; B31872.
CC HSSP: P29373; ICBS.
CC InterPro: IPR000463; Fatty_acid_BP.
CC InterPro: IPR000566; Lipocalin_cytfabp.
CC Pfam: PF00061; Lipocalin; 1.

```

```

DR PROSITE; PS00214; FAPB; 1.
KW Vitamin A; Transport.
SQ NON TER 36
SQ SEQUENCE 36 AA; 3997 MW; 2BC1C154B2FD9904 CRC64;
OY 1 WKMKPRNLEE 11
Db 7 WKMKSENFEF 17
OY 1 WKMKPRNLEE 11
Db 7 WKMKSENFEF 17

RESULT 12
RET3_XENIA STANDARD; PRT; 137 AA.
AC P50568;
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Retinoic acid-binding protein, cellular (CRABP).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95034308; PubMed=7947321;
RA Ho L., Mercola M., Gudas L.J.;
RT "Xenopus laevis cellular retinoic acid-binding protein: temporal and
RT spatial expression pattern during early embryogenesis.";
RL Mech. Dev. 47:53-64(1994).
CC -1- FUNCTION: CYTOSOLIC CRABPS MAY REGULATE THE ACCESS OF RETINOIC
CC ACID TO THE NUCLEAR RETINOIC ACID RECEPTORS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE PABP/E2/CRBP/CRABP FAMILY OF
CC TRANSPORTERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S74933; AB032580.1; -.
DR HSSP; P29373; ICBS.
DR InterPro; IPR000463; Fatty_acid_BP.
DR InterPro; IPR000566; Lipocin_cytfABP.
DR Pfam; PF00061; lipocalin; 1.
DR PRINTS; PR00178; FATTYACIDBP.
DR PROSITE; PS00214; FAPB; 1.
KW Vitamin A; Transport.
KW INIT MET 0
SQ SEQUENCE 137 AA; 15524 MW; 2193DDE7B095EC4A CRC64;
BY SIMILARITY.

Query Match 46.5%; Score 40; DB 1; Length 36;
Best Local Similarity 63.6%; Pred. No. 1.6;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 WKMKPRNLEE 11
Db 7 WKMKSENFEF 17

RESULT 13
118B_HUMAN STANDARD; PRT; 197 AA.
AC O95938; O96027; O95993; Q9UBR7;
DT 30-MAY-2000 (Rel. 39; Created)

```

```

DT      30-MAY-2000 (Rel. 39, Last sequence update)
DR      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Interleukin-18 binding protein precursor (IL-18BP).
GN      IL18BP.
OS      Homo sapiens (Human) .
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A., SEQUENCE OF 29-68, FUNCTION, ALTERNATIVE SPLICING,
AP      AND TISSUE SPECIFICITY.
RX      MEDLINE=99146382; PubMed=10023777;
RA      Novick D., Kim S.-H., Fantuzzi G., Reznikov L.L., Dinarello C.A.,
RT      Rubinstein M.;
RL      "Interleukin-18 binding protein: a novel modulator of the Th1 cytokine
response.";
RM      Immunity 10:127-136(1999) .
RN      [2]
RP      SEQUENCE FROM N.A. (ISOFORM A), SEQUENCE OF 29-33; 37-43; 46-52;
AP      55-58; 61-72; 88-93; 105-142 AND 145-167, AND TISSUE SPECIFICITY.
RX      MEDLINE=99192308; PubMed=10094485;
RA      Alawa Y., Atila K., Tanlali M., Toriogoe K., Mori T., Niehida Y.,
RS      Ushio S., Nukada Y.T., Tanimoto T., Ikegami H., Ikeda M., Kuramoto M.;
RT      "Cloning and expression of interleukin-18 binding protein.";
RL      FEBS Lett. 445:338-342(1999) .
RN      [3]
RP      SEQUENCE FROM N.A. (ISOFORM A).
RX      MEDLINE=99263157; PubMed=10329540;
RA      Xiang Y., Moss B.;
RT      "Identification of human and mouse homologs of the MCS1L-53L-54L
family of secreted glycoproteins encoded by the Moluscum contagiosum
porvirus.";
RL      Virology 257:297-302(1999) .
RN      -1-
FP      FUNCTION: Binds to IL-18 and inhibits its activity. Functions as
an inhibitor of the early TH1 cytokine response.
CC      -1- SUBCELLULAR LOCATION: Secreted (Potential) .
CC      -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: A/I-L-18BA, B/I-L-18BB AND C/I-L-
18BPC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC      -1- TISSUE SPECIFICITY: Strongly expressed in heart, lung, placenta
and spleen.
CC      -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC      -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no way commercially modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announcement/ or send an email to license@ebi.ac.uk).
CC      -----
DR      EMBL; AF110798; AACD17187.1; -
DR      EMBL; AF110798; AACD17188.1; -
DR      EMBL; AF110798; AACD17189.1; -
DR      EMBL; AF110799; AACD17190.1; -
DR      EMBL; AF110800; AACD17191.1; -
DR      EMBL; AF110801; AACD17192.1; -
DR      EMBL; AB019504; BAAT6374.1; -
DR      EMBL; AF112906; AAA41051.1; ALT_INIT.
DR      Geneweb; HGNC:5987; IL18BP.
DR      MIM; 604113; -
DR      InterPro; IPR003006; Ig_MHC.
DR      Pfam; PF00047; Ig_1.
KW      Immunoglobulin domain; Glycoprotein; signal; Alternative splicing.
FT      SIGNAL                                1..28
FT      CHAIN                                  29..197
FT      DOMAIN                                77..155
FT      DISULFID                               84..148
FT      CARBOHYD                              77..77
FT      CARBOHYD                              92..92
FT      CARBOHYD                             101..101
FT      CARBOHYD                             145..145
FT      VARIANT                               168..197
FT      VARIANT                               168..197

```

```

PT  VARSPPLIC 77 113 PPTQALPSSSSPQOOG (IN ISOFORM A).
PT  VARSPPLIC 77 113 NGTSLSCVACSRFPNPSILYMGNGSFIEHLPRLM ->
PT  VARSPPLIC 114 197 SNAEGNLAIPRSPALQPGQSTAAGRILSTGPAAAGP (IN
PT  VARSPPLIC 114 197 ISOFORM B).
SQ  SEQUENCE 197 AA; 21698 MW; 21698 MW; FID626AB2B285E1 CRC64;

Query Match
Best Local Similarity 46.7%; Pred. No. 11;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 WKMKPRNLEEDYL 15
Db 167 WVRSPRGLQOEEL 181

RESULT 14
SYR ARCFU STANDARD; PRT; 549 AA.
AC 029368;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine-tRNA ligase) (ArgRS).
GN ARGS OR AF0894.
OS Archaeoglobus fulgidus.
OC Archaeoglobaceae; Archaeoglobus.
NCBI_TaxID=2234;
RN 11)
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.B.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrtides N.C.,
RA Fleischmann D., Kachanubush J., Lee N.H., Sutton G.G., Gill S.,
RA Peterson E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Corton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason J.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -1- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
CC diphosphate + L-arginyl-tRNA(Arg).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE001042; AAB90346.1; -.
DR TIGR: AF0894; -.
DR InterPro: IPR001278; Arg_tRNA-synt_1c.
DR InterPro: IPR005148; N_Arg_1.
DR InterPro: IPR001412; tRNA-synt_1.
DR Pfam: PF00750; tRNA-synt_1d; 1.
DR Pfam: PF03485; N-Arg; 1.
DR PRINTS: PRO1038; TRNASYNTHARG.
DR TIGRfam: TIGR00456; ARGS; 1.
DR PROSITE: PS00178; AA_tRNA_LIGASE_I; FALSE_NEG.
KW Aminocyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SITE 113 123 "HIGH" REGION.

```

```

SQ SEQUENCE 549 AA; 62861 MW; 39DF41CC0B9AD210 CRC64;

Query Match
Best Local Similarity 46.5%; Score 40; DB 1; Length 549;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 WKMKPRNLEED 13
Db 377 WRLISRDMEDE 389

RESULT 15
GPT1_CANAL
ID GPT1_CANAL STANDARD; PRT; 553 AA.
AC 074248;
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE Putative polyamine transporter.
GN GPT1.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mtosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN 11)
RP SEQUENCE FROM N.A.
RA McNamee M.D., Gorman J.A., Buckley H.R.;
RT "Isolation and sequence of the GPT1 gene of Candida albicans, encoding
RT a putative polyamine transporter.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF080132; AAC31569.1; -.
DR InterPro: IPR002293; AA/tel_permase1.
DR InterPro: IPR004840; AAC_permasee.
DR InterPro: IPR004841; Permease.
DR Pfam: PF00324; aa_permeases; 1.
DR PROSITE: PS00218; AMINO_ACID_PERMEASE_1; FALSE_NEG.
KW Transport; Transmembrane.
FT TRANSMEM 43 63 POTENTIAL.
FT TRANSMEM 68 88 POTENTIAL.
FT TRANSMEM 133 153 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
FT TRANSMEM 190 210 POTENTIAL.
FT TRANSMEM 240 260 POTENTIAL.
FT TRANSMEM 280 300 POTENTIAL.
FT TRANSMEM 333 353 POTENTIAL.
FT TRANSMEM 397 417 POTENTIAL.
FT TRANSMEM 453 473 POTENTIAL.
FT TRANSMEM 485 505 POTENTIAL.
SQ SEQUENCE 553 AA; 60633 MW; 197DACB76C5FBC9E CRC64;

Query Match
Best Local Similarity 46.2%; Score 40; DB 1; Length 553;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 WKMKPRNLEED 13
Db 508 WYVGPRNISEED 520

Search completed: March 26, 2003, 16:45:03
Job time : 5.36047 secs

```



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 16:44:15 ; Search time 8.54651 Seconds  
(without alignments)  
168.726 Million cell updates/sec

Title: US-10-010-667A-2\_COPY\_14\_28

Perfect score: 86  
Sequence: 1 WKMKPRRLREDDYL 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by change to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45.5	52.9	237	2	T50202 conserved hypothet
2	45	52.3	375	1	NRECD ribonuclease III (
3	45	52.3	375	2	A99943 RNase D, processes
4	45	52.3	375	2	E85791 RNase D, processes
5	44	51.2	532	2	T02539 hypothetical prote
6	43	50.0	312	2	T15371 hypothetical prote
7	43	50.0	539	2	T39150 probable heat shoc
8	42	48.8	124	2	AH1754 bacteriophage prot
9	42	48.8	384	2	T49020 hypothetical prote
10	42	48.8	416	2	D96692 hypothetical prote
11	42	48.8	442	2	D36718 dihydrolipeamide S
12	42	48.8	451	2	T49021 hypothetical prote
13	42	48.8	847	2	JC4836 alpha-glucuronidas
14	42	48.8	1832	2	AC2594 glutamate synthase
15	42	48.8	1858	2	D97376 ABC-type transport
16	41	47.7	260	2	S75010 hypothetical prote
17	41	47.7	271	2	G84550 NADH oxidase-relat
18	41	47.7	403	2	G75287 hypothetical prote
19	41	47.7	580	2	T21493 transcription fact
20	41	47.7	1187	2	T46637 probable heat shoc
21	40.5	47.1	639	2	F71258 hypothetical prote
22	40.5	47.1	658	2	S76909 immediate-early pr
23	40.5	47.1	1415	1	EDBERGA retinoid acid-bind
24	40	46.5	36	2	B31872 xCRABP - African c
25	40	46.5	138	2	T51265 hypothetical prote
26	40	46.5	361	2	T49337 hypothetical prote
27	40	46.5	430	2	T23899 hypothetical prote
28	40	46.5	474	2	T00943 hypothetical prote
29	40	46.5	526	2	F85086 hypothetical prote

30	40	46.5	549	1	E69361 arginyl-tRNA synth
31	40	46.5	884	2	H83322 hypothetical prote
32	40	46.5	961	2	T01167 hypothetical prote
33	40	46.5	1268	2	T18955 hypothetical prote
34	39	45.3	90	2	H90665 probable Ogr fam1
35	39	45.3	90	2	C85516 probable activator
36	39	45.3	371	2	C64499 isocitrate dehydro
37	39	45.3	426	2	E83981 pyruvate dehydroge
38	39	45.3	426	2	B96519 protein T2E6.2 (Im
39	39	45.3	449	2	G89841 hypothetical prote
40	39	45.3	479	2	B69764 transcription regu
41	39	45.3	512	2	T23035 hypothetical prote
42	39	45.3	652	2	B84568 probable calmoduli
43	39	45.3	713	2	JC5870 poly(beta-D-mannur
44	39	45.3	976	2	A97104 Zn-dependent metal
45	39	45.3	1019	2	T00117 dve protein - fru1

## ALIGNMENTS

## RESULT 1

T50202 conserved hypothetical protein SPAC25B8.15c [imported] - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 09-Jun-2000  
C:Accession: T50202  
R:Barrell, B.G.; Rajandream, M.A.; McDougall, R.C.; McLean, J.; Harris, D. submitted to the EMBL Data Library, November 1999  
A:Reference number: 225045  
A:Accession: T50202  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-237 <BAR>  
A:Cross-references: EMBL:AL13225; PDB:CA61781.1; GSPDB:GN00066; SPDB:SPAC25B8.15c  
A:Experimental source: strain 972h(-); cosmid c25B8  
C:Genetics:  
A:Gene: SPDB:SPAC25B8.15c  
A:Map position: 1

Query Match 52.9%; Score 45.5; DB 2; Length 237;  
Best Local Similarity 66.7%; Pred. No. 4.3;  
Matches 10; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 2 KMKPRRLREDDYL 15  
DB 220 KMKPRRLREDDYL 234

## RESULT 2

NRECD ribonuclease III (EC 3.1.26.3) rnd - Escherichia coli (strain K-12)  
N:Alternate names: ribonuclease D  
C:Species: Escherichia coli  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 01-Mar-2002  
C:Accession: S01223; A30431; S41590; D64941; S42849; D21915  
N:Zhang, J.; Deutscher, M.P.  
A:Title: Escherichia coli RNase D: sequencing of the rnd structural gene and purification  
A:Reference number: S01223; MUID:88289400; PMID:3041371  
A:Accession: S01223  
A:Molecule type: DNA  
A:Residues: 1-375 <ZHA>  
A:Cross-references: EMBL:X07055; NID:942770; PDB:CAA30098.1; PID:G561215  
A:Accession: A30431  
A:Molecule type: protein  
A:Residues: 1-6 <ZH2>  
R:Puida, M.; Heinz, E.; Wolter, F.P.  
Mol. Gen. Genet. 242, 241-249, 1994  
A:Title: The fnd gene of Escherichia coli K12 is located close to rnd at 39.6 min of the  
A:Reference number: S41588; MUID:94150456; PMID:8107670  
A:Accession: S41590  
A:Molecule type: DNA

A:Residues: 1-38 <FUL>  
 A:Cross-references: GB:X70994; NID:g33478; PIDN:CAA50322.1; PID:g581071  
 A:Experimental source: strain K12  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co-  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of *Escherichia coli* K-12.  
 A:Reference number: A64720; PMID:9278503  
 A:Accession: D64941  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-375 <BLAT>  
 A:Cross-references: GB:AE00274; GB:U00096; NID:g1788089; PIDN:AAC74874.1; PID:g1788105;  
 A:Experimental source: strain K-12, substrain MG1655  
 R:Paula, M.  
 submitted to the EMBL Data Library, February 1993  
 A:Reference number: S42848  
 A:Accession: S42849  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 'L', 2-38 <FU2>  
 A:Cross-references: EMBL:X70994  
 C:Genetics:  
 A:Gene: rnd  
 A:Map position: 40 min  
 A:Start codon: TTG  
 C:Superfamily: ribonuclease D  
 C:Keywords: exonuclease; hydrolase  
 F1-375/Product: ribonuclease D #stratus experimental <MAT>

Query Match 52.3%; Score 45; DB 1; Length 375;  
 Best Local Similarity 63.6%; Pred. No. 8.5;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 WKMKPRNLEE 11  
 ||:|:|:|  
 Db 342 WKLPQNLP 352

RESULT 3  
 A:Residues: 1-375 <RND>  
 A:Cross-references: EMBL:AC004684; NID:g3236234; PID:g3236255  
 A:Experimental source: cultivar Columbia  
 C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
 C:Accession: A99943  
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
 gatawara, N.; Yasunaga, T.; Kohara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and geno-  
 A:Reference number: A99629; PMID:11258796  
 A:Accession: A99943  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-375 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BAH35936.1; PID:g13361980; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RIMD 0509952  
 C:Genetics:  
 A:Gene: ECE2513  
 C:Superfamily: ribonuclease D

Query Match 52.3%; Score 45; DB 2; Length 375;  
 Best Local Similarity 63.6%; Pred. No. 8.5;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 WKMKPRNLEE 11  
 ||:|:|:|  
 Db 342 WKLPQNLP 352

RESULT 4  
 E85791  
 RNase D, processes tRNA precursor (imported) - *Escherichia coli* (strain O157:H7, substra-  
 C:Species: *Escherichia coli*

C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: E85791  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew,  
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousta, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
 A:Reference number: A85480; PMID:11074935; PMID:11206551  
 A:Accession: E85791  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-375 <STO>  
 A:Cross-references: GB:AE005174; NID:g12515844; PIDN:AA656793.1; GSPDB:GN00145; UMGF:Z284  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: rnd  
 C:Superfamily: ribonuclease D

Query Match 52.3%; Score 45; DB 2; Length 375;  
 Best Local Similarity 63.6%; Pred. No. 8.5;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 WKMKPRNLEE 11  
 ||:|:|:|  
 Db 342 WKLPQNLP 352

RESULT 5  
 T02539  
 hypothetical protein Ac2g37730 (imported) - *Arabidopsis thaliana*  
 N:Alternate names: hypothetical protein F13M22.23  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C:Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 16-Feb-2001  
 C:Accession: T02539; D84796  
 R:Roundley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.  
 submitted to the EMBL Data Library, June 1998  
 A:Description: *Arabidopsis thaliana* chromosome II BAC F13M22 genomic sequence.  
 A:Reference number: Z14677  
 A:Accession: T02539  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-532 <ROU>  
 A:Cross-references: EMBL:AC004684; NID:g3236234; PID:g3236255  
 A:Experimental source: cultivar Columbia  
 R:Lin, X.; Kaul, S.; Roundley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Talion, L.;  
 eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
 A:Reference number: A84420; PMID:10617197  
 A:Accession: D84796  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-532 <STO>  
 A:Cross-references: GB:AE002093; NID:g3236255; PIDN:AAC3643.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: Ac2g37730; F13M22.23  
 A:Map position: 2  
 A:Introns: 165/2; 295/3; 478/3  
 C:Superfamily: *Arabidopsis* hypothetical protein F13M22.23

Query Match 51.2%; Score 44; DB 2; Length 532;  
 Best Local Similarity 55.6%; Pred. No. 18;  
 Matches 10; Conservative 0; Mismatches 2; Indels 6; Gaps 1;

QY 1 WKMKPRN-----NLEED 12  
 ||:|:|:|  
 Db 476 WKMAPRQCEIVNSSED 493

RESULT 6  
 T15371  
 hypothetical protein C01F1.2 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*

C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C/Accession: T15371  
R:Johnson, D.  
Submitted to the EMBL Data Library, May 1996  
A/Description: The sequence of *C. elegans* cosmid C01F1.  
A/Reference number: Z18338  
A/Accession: T15371  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-312 <J0H>  
A/Cross-references: EMBL:U58761; NID:g1330391; PID:g1330396; PIDN:AA800716.1; GSPDB:GN00  
A/Experimental source: strain Bristol NZ; clone C01F1  
A/Genetic: CESP:C01F1.2  
A/Map position: 2  
A/Introns: 36/2; 271/1

Query Match 50.0%; Score 43; DB 2; Length 312;  
Best Local Similarity 63.6%; Pred. No. 15;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 PRNRLEDDYL 15  
DB 250 PRNRQEDDYI 260

RESULT 7  
T39150  
Probable heat shock transcription factor - fission yeast (*Schizosaccharomyces pombe*)  
C/Species: *Schizosaccharomyces pombe*  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C/Accession: T39150  
R:Oliver, K.; Harris, D.; Bartrell, B.G.; Rajandream, M.A.; Wood, V.  
Submitted to the EMBL Data Library, September 1997  
A/Reference number: Z21748  
A/Accession: T39150  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-539 <OLY>  
A/Cross-references: EMBL:Z99168; PIDN:CA81301.1; GSPDB:GN00066; SPDB:SPAC8C9.14  
A/Experimental source: strain 97zh-; cosmid c8C9  
C/Genetic:  
A/Genetic: SPDB:SPAC8C9.14  
A/Map position: 1  
A/Introns: 10/3; 40/3; 67/2; 86/3

Query Match 50.0%; Score 43; DB 2; Length 539;  
Best Local Similarity 60.0%; Pred. No. 27;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 WKMKPRNRLEDDYL 15  
DB 364 WKRPRLVDEDEL 378

RESULT 8  
AH1754  
bacteriophage protein homolog l1n2581 [imported] - *Listeria innocua* (strain Clijp11262)  
C/Species: *Listeria innocua*  
C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C/Accession: AH1754  
R:Glaser, P.; Frangoul, L.; Buchleser, C.; Amend, A.; Baquero, F.; Barche, P.; Bloeker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feiht, H.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A/Authors: Krefl, J.; Kuhn, M.; Kunst, F.; Kuzepkat, G.; Madueno, E.; Maltounam, A.; Maok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A./Title: Comparative genomics of *Listeria* species.  
A/Reference number: AB1077; MUID:2157279; PMID:11679669  
A/Accession: AH1754  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-124 <GLA>

A/Cross-references: GB:AL592022; PIDN:CA97808.1; PID:g16415103; GSPDB:GN00178  
A/Experimental source: strain Clijp11262  
C/Genetic:  
A/Genetic: l1n2581

Query Match 48.8%; Score 42; DB 2; Length 124;  
Best Local Similarity 46.7%; Pred. No. 8; 6;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 WKMKPRNRLEDDYL 15  
DB 15 WKMKKXILRDEYL 29

RESULT 9  
T49020  
Hypothetical protein F3C22.70 - *Arabidopsis thaliana* (fragment)  
C/Species: *Arabidopsis thaliana* (mouse-ear cross)  
C/Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C/Accession: T49020  
R:Purcell, B.; Maury, D.; Goffeau, A.; Boutry, M.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Me submitted to the Protein Sequence Database, April 2000  
A/Reference number: Z25013  
A/Accession: T49020  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-384 <PUR>  
A/Cross-references: EMBL:AL353912; GSPDB:GN00061; ATSP:F3C22.70  
A/Experimental source: cultivar Columbia; BAC clone F3C22  
C/Genetic:  
A/Genetic: ATSP:F3C22.70  
A/Map position: 3  
A/Introns: 61/1; 252/3; 281/3

Query Match 48.8%; Score 42; DB 2; Length 384;  
Best Local Similarity 42.9%; Pred. No. 28;  
Matches 6; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 WKMKPRNRLEDDYL 14  
DB 46 WKLVPLRFDSDY 59

RESULT 10  
D96692  
Hypothetical protein T1217.9 [imported] - *Arabidopsis thaliana*  
C/Species: *Arabidopsis thaliana* (mouse-ear cross)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C/Accession: D96692  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizart, L.  
Nature 408, 816-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luers, J.S.; Maiti, R.; Marshall, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Iker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
A/Reference number: AB6141; MUID:21016719; PMID:11130712  
A/Accession: D96692  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-416 <STO>  
A/Cross-references: GB:AE005173; NID:g11054573; PIDN:AA627848.1; GSPDB:GN00141  
C/Genetic:  
A/Genetic: T1217.9  
A/Map position: 1

Query Match 48.8%; Score 42; DB 2; Length 416;  
Best Local Similarity 50.0%; Pred. No. 31;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 WKMKPRRNLEDDY 14  
Db 40 WKVPRRLDEADP 53

## RESULT 11

D36718  
dihydrolipoamide S-acyltransferase (EC 2.3.1.12) precursor - *Bacillus subtilis*  
N:Alternate names: pyruvate dehydrogenase complex, E2 component  
C:Species: *Bacillus subtilis*  
C>Date: 12-Apr-1991 #sequence\_revision 12-Apr-1991 #text\_change 20-Jun-2000  
R:Hemila, H.; Palva, A.; Paulin, L.; Arvidsson, S.; Palva, I.  
J. Bacteriol. 172, 5052-5063, 1990  
A:Title: Secretory S complex of *Bacillus subtilis*: sequence analysis and identity to pyruvate dehydrogenase E2 component  
A:Reference number: A36718; PMID:90368558; PMID:1697575  
A:Accession: D36718  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-442 <HEM>  
A:Cross-references: GB:M57435; GB:M31542; NID:G143375; PIDN:AAA62683.1; PID:G143379  
R:Kunze, F.; Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berth  
C.: Bron, S.; Brouillet, S.; Bruneau, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
A.: Enrich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabel, C.; Ferrati, B.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallazzi, A.; Gall  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holzapfel, S.; Hosono, S.; Huillo, M.F.  
Koetter, P.; Konings, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, S.; Lardinois  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maeda, S.; Maue  
Y., M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portelle  
Rieger, M.; Rivolta, C.; Rocha, B.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon  
A:Authors: Schleicher, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Ser  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; PMID:8044033; PMID:9384377  
A:Accession: B69674  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-442 <KUN>  
A:Cross-references: GB:299111; GB:AL009126; NID:G2633699; PIDN:CAB13333.1; PID:G2633831  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: pdhC  
C:Superfamily: dihydrolipoamide acetyltransferase; lipoyl/biotin-binding homology  
C:Keywords: acyltransferase; coenzyme A; oxidoreductase  
F:4-77/Domain: lipoyl/biotin-binding homology <LPB>  
F:413.417/Active site: His, Asp #status predicted

## Query Match

Best Local Similarity 48.8%; Score 42; DB 2; Length 442;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 WKMKPRRNLEDDY 15  
Db 22 WKVPRRLDEADP 36

## RESULT 12

T49021  
hypothetical protein F3C22.80 - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C>Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C:Accession: T49021  
R:Purnelle, B.; Masny, D.; Goffeau, A.; Bouty, M.; Mewes, H.W.; Rudd, S.; Lemcke, K.; M  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z25013  
A:Accession: T49021  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-451 <PUR>  
A:Cross-references: EMBL:ALJ53912; GSPDB:GN00061; ATSP:F3C22.80

A:Experimental source: cultivar Columbia; BAC clone F3C22  
C:Genetics:  
A:Gene: ATSP:F3C22.80  
A:Map position: 3  
A:Introns: 4/2; 294/3; 339/3

Query Match 48.8%; Score 42; DB 2; Length 451;  
Best Local Similarity 42.9%; Pred. No. 33;  
Matches 6; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 WKMKPRRNLEDDY 14  
Db 52 WKVPRRLDEADP 65

## RESULT 13

JC4836  
alpha-glucuronidase (EC 3.2.1.-) precursor - fungus (*Trichoderma reesei*)  
N:Alternate names: GLR1  
C:Species: *Trichoderma reesei*  
C>Date: 15-Aug-1996 #sequence\_revision 15-Oct-1996 #text\_change 17-Mar-1999  
C:Accession: JC4836  
R:Margolles-Lark, E.; Saloheimo, M.; Silke-aho, M.; Penttilae, M.  
Gene 172, 171-172, 1996  
A:Title: The alpha-glucuronidase-encoding gene of *Trichoderma reesei*.  
A:Reference number: JC4836; PMID:96257277; PMID:8654984  
A:Accession: JC4836  
A:Molecule type: mRNA  
A:Residues: 1-847 <MAR>  
A:Cross-references: EMBL:268706; NID:G1419337; PID:G218512; PID:G1419338  
C:Comment: This enzyme releases glucuronic acid attached to xylose units of xylan.  
C:Genetics:  
A:Gene: glr1  
C:Keywords: glycosidase; hydrolase  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-847/Product: alpha-glucuronidase #status predicted <MAR>

Query Match 48.8%; Score 42; DB 2; Length 847;  
Best Local Similarity 53.8%; Pred. No. 64;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 WKMKPRRNLEDDY 14  
Db 108 KLSPKLTKEDGY 120

## RESULT 14

AC2594  
glutamate synthase large subunit gltB [imported] - *Agrobacterium tumefaciens* (strain C58,  
C:Species: *Agrobacterium tumefaciens*  
C>Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C:Accession: AC2594  
R:Wood, D.W.; Seubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, L  
erage, G.; Gillet, W.; Grant, C.; Gunther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AC2594  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1832 <KUN>  
A:Cross-references: GB:AB008668; PIDN:AAI41169.1; PID:G17738468; GSPDB:GN00186  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: gltB  
A:Map position: circular chromosome

Query Match 48.8%; Score 42; DB 2; Length 1832;  
Best Local Similarity 53.8%; Pred. No. 1,4e+02;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 MKPRRLLEDDYL 15  
::||:||||  
DB 356 LRPLRTVERTDYL 368

## RESULT 15

D97376  
hypothetical protein AGR\_C\_235 [imported] - Agrobacterium tumefaciens (strain C58, Cerec  
C:Species: Agrobacterium tumefaciens  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C:Accession: D97376  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughly, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A:Reference number: A97359; PMID:11743194  
A:Accession: D97376  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-1858 <KUR>  
A:Cross-references: GB:AE007869; PIDN:AAK85965.1; PID:g15155024; GSPDB:GN00169  
C:Genetics:  
A:Gene: AGR\_C\_235  
A:Map position: circular chromosome

Query Match 48.8%; Score 42; DB 2; Length 1858;  
Best Local Similarity 53.8%; Pred. No. 1.5e+02;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 3 MKPRRLLEDDYL 15  
::||:||||  
DB 382 LRPLRTVERTDYL 394

Search completed: March 26, 2003, 16:50:36  
Job time : 12.5465 secs



GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 16:49:57 ; Search time 10.1163 Seconds  
(without alignments)  
87.073 Million cell updates/sec

Title: US-10-010-667a-21

Perfect score: 93

Sequence: 1 WIDIKQFWYTPPTF 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 237916 seqs, 58723674 residues

Total number of hits satisfying chosen parameters: 237916

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US05\_PUBCOMB pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	93	100.0	15	9 US-10-011-095-21	Sequence 21, Appl
2	93	100.0	15	9 US-10-010-667a-21	Sequence 21, Appl
3	93	100.0	267	10 US-09-747-835A-50	Sequence 50, Appl
4	93	100.0	267	10 US-09-747-835A-51	Sequence 51, Appl
5	93	100.0	339	9 US-10-012-896-879	Sequence 879, Appl
6	93	100.0	339	9 US-09-802-520-11	Sequence 11, Appl
7	93	100.0	339	9 US-09-895-793-879	Sequence 879, Appl
8	93	100.0	339	9 US-09-895-814-879	Sequence 879, Appl
9	93	100.0	339	9 US-10-011-095-2	Sequence 2, Appl
10	93	100.0	339	9 US-10-010-667a-2	Sequence 2, Appl
11	93	100.0	339	10 US-09-759-143-879	Sequence 879, Appl
12	93	100.0	339	10 US-09-780-669-879	Sequence 879, Appl
13	93	100.0	339	10 US-09-822-827-879	Sequence 879, Appl
14	93	100.0	268	10 US-09-854-816-39	Sequence 39, Appl
15	93	100.0	268	10 US-09-854-816-33	Sequence 33, Appl
16	93	100.0	268	10 US-09-815-242-5492	Sequence 5492, Appl
17	93	100.0	568	10 US-09-815-242-12629	Sequence 12629, Appl
18	93	100.0	147	10 US-09-873-134-2	Sequence 2, Appl
19	93	100.0	351	9 US-10-026-741-47	Sequence 47, Appl

20	43	46.2	873	1 US-08-911-824-61	Sequence 61, Appl
21	43	46.2	877	9 US-10-026-741-102	Sequence 102, Appl
22	42	45.2	268	10 US-09-854-816-26	Sequence 26, Appl
23	42	45.2	268	10 US-09-854-816-68	Sequence 68, Appl
24	42	45.2	268	10 US-09-854-816-71	Sequence 71, Appl
25	42	45.2	459	9 US-10-007-527A-21	Sequence 21, Appl
26	42	45.2	619	10 US-09-891-609-4	Sequence 2, Appl
27	42	45.2	646	10 US-09-891-609-2	Sequence 2, Appl
28	42	45.2	847	10 US-09-476-242-2	Sequence 2, Appl
29	41	44.1	243	10 US-09-854-816-73	Sequence 73, Appl
30	41	44.1	268	10 US-09-854-816-8	Sequence 8, Appl
31	41	44.1	268	10 US-09-854-816-13	Sequence 13, Appl
32	41	44.1	268	10 US-09-854-816-14	Sequence 14, Appl
33	41	44.1	268	10 US-09-854-816-70	Sequence 70, Appl
34	41	44.1	268	10 US-09-854-816-72	Sequence 72, Appl
35	41	44.1	269	10 US-09-854-816-21	Sequence 21, Appl
36	41	44.1	269	10 US-09-854-816-22	Sequence 22, Appl
37	41	44.1	269	10 US-09-854-816-25	Sequence 25, Appl
38	41	44.1	269	10 US-09-854-816-37	Sequence 37, Appl
39	41	44.1	270	10 US-09-854-816-67	Sequence 67, Appl
40	41	44.1	351	9 US-09-886-156-46	Sequence 46, Appl
41	41	44.1	351	9 US-09-886-156-45	Sequence 46, Appl
42	41	44.1	351	9 US-09-886-149-46	Sequence 46, Appl
43	41	44.1	351	9 US-09-886-159-46	Sequence 46, Appl
44	41	44.1	625	9 US-10-032-162-17	Sequence 17, Appl
45	41	44.1	655	10 US-09-725-735A-12	Sequence 12, Appl

#### ALIGNMENTS

RESULT 1  
US-10-011-095-21  
Sequence 21, Application US/10011095  
Publication No. US20030045682A1  
GENERAL INFORMATION:  
APPLICANT: Afar, Daniel  
APPLICANT: Hubert, Rene S.  
APPLICANT: Leong, Kahan  
APPLICANT: Raitano, Arthur B.  
APPLICANT: Saitan, Douglas C.  
APPLICANT: Mitchell, Steve Chapell  
TITLE OF INVENTION: ANTIBODIES IMMUNOSPECIFIC FOR STEAP1 (AS AMENDED)  
FILE REFERENCE: 511582001610  
CURRENT APPLICATION NUMBER: US/10/011,095  
CURRENT FILING DATE: 2001-12-06  
PRIOR APPLICATION NUMBER: 09/323,873  
PRIOR FILING DATE: 1999-06-01  
PRIOR APPLICATION NUMBER: 60/087,520  
PRIOR FILING DATE: 1998-06-01  
PRIOR APPLICATION NUMBER: 60/091,183  
PRIOR FILING DATE: 1998-06-30  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 21  
LENGTH: 15  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: STEAP-1 PEPTIDE  
US-10-011-095-21

Query Match 100.0%; Score 93; DB 9; Length 15;  
Best Local Similarity 100.0%; Pred. No. 4.5e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WIDIKQFWYTPPTF 15  
|||  
Db 1 WIDIKQFWYTPPTF 15

RESULT 2  
US-10-010-667a-21

```
/ Sequence 21, Application US/10010667A
/ Publication No. US20030055217A1
/ GENERAL INFORMATION:
/ APPLICANT: Afar, Daniel
/ APPLICANT: Hubert, Rene S.
/ APPLICANT: Leong, Kahen
/ APPLICANT: Raitano, Arthur B.
/ APPLICANT: Saffran, Douglas C.
/ APPLICANT: Mitchell, Steve Chappell
/ TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
/ FILE REFERENCE: 511582001601
/ CURRENT APPLICATION NUMBER: US/10/010,667A
/ CURRENT FILING DATE: 2001-12-06
/ PRIOR APPLICATION NUMBER: 09/323,873
/ PRIOR FILING DATE: 1999-06-01
/ PRIOR APPLICATION NUMBER: 60/087,520
/ PRIOR FILING DATE: 1998-06-01
/ PRIOR APPLICATION NUMBER: 60/091,183
/ PRIOR FILING DATE: 1998-06-30
/ NUMBER OF SEQ ID NOS: 32
/ SOFTWARE: PastSeq for Windows Version 4.0
/ SEQ ID NO 21
/ LENGTH: 15
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: STEAP-1 PEPTIDE
/ US-10-010-667a-21
```

```
Query Match 100.0%; Score 93; DB 9; Length 15;
Best Local Similarity 100.0%; Pred. No. 4.5e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 WIDIKQFWYTPPTF 15
Db 1 WIDIKQFWYTPPTF 15
```

```
RESULT 3
US-09-747-835A-50
/ Sequence 50, Application US/09747835A
/ Patent No. US20020146692A1
/ GENERAL INFORMATION:
/ APPLICANT: Yamazaki, Victoria
/ APPLICANT: Tang, Y. Tom
/ APPLICANT: Liu, Chenghua
/ APPLICANT: Zhou, Ping
/ APPLICANT: Wang, Dunrui
/ APPLICANT: Zhang, Jie
/ APPLICANT: Ren, Feiyan
/ APPLICANT: Asundi, Vinod
/ APPLICANT: Dmanac, Radoje T
/ TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-LIKE
/ FILE REFERENCE: HYS-37CIP
/ CURRENT APPLICATION NUMBER: US/09/747,835A
/ CURRENT FILING DATE: 2002-03-08
/ PRIOR APPLICATION NUMBER: US 09/729,739
/ PRIOR FILING DATE: 2000-12-04
/ PRIOR APPLICATION NUMBER: US 09/653,450
/ PRIOR FILING DATE: 2000-08-31
/ PRIOR APPLICATION NUMBER: US 09/620,312
/ PRIOR FILING DATE: 2000-07-19
/ PRIOR APPLICATION NUMBER: US 09/598,042
/ PRIOR FILING DATE: 2000-06-20
/ PRIOR APPLICATION NUMBER: US 09/552,317
/ PRIOR FILING DATE: 2000-04-25
/ PRIOR APPLICATION NUMBER: US 09/488,725
/ PRIOR FILING DATE: 2000-01-21
/ NUMBER OF SEQ ID NOS: 63
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 50
```

```
/ LENGTH: 267
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ US-09-747-835A-50
```

```
Query Match 100.0%; Score 93; DB 10; Length 267;
Best Local Similarity 100.0%; Pred. No. 8.2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 WIDIKQFWYTPPTF 15
Db 211 WIDIKQFWYTPPTF 225
```

```
RESULT 4
US-09-747-835A-51
/ Sequence 51, Application US/09747835A
/ Patent No. US20020146692A1
/ GENERAL INFORMATION:
/ APPLICANT: Yamazaki, Victoria
/ APPLICANT: Tang, Y. Tom
/ APPLICANT: Liu, Chenghua
/ APPLICANT: Zhou, Ping
/ APPLICANT: Wang, Dunrui
/ APPLICANT: Zhang, Jie
/ APPLICANT: Ren, Feiyan
/ APPLICANT: Asundi, Vinod
/ APPLICANT: Dmanac, Radoje T
/ TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-LIKE
/ FILE REFERENCE: HYS-37CIP
/ CURRENT APPLICATION NUMBER: US/09/747,835A
/ CURRENT FILING DATE: 2002-03-08
/ PRIOR APPLICATION NUMBER: US 09/729,739
/ PRIOR FILING DATE: 2000-12-04
/ PRIOR APPLICATION NUMBER: US 09/653,450
/ PRIOR FILING DATE: 2000-08-31
/ PRIOR APPLICATION NUMBER: US 09/620,312
/ PRIOR FILING DATE: 2000-07-19
/ PRIOR APPLICATION NUMBER: US 09/598,042
/ PRIOR FILING DATE: 2000-06-20
/ PRIOR APPLICATION NUMBER: US 09/552,317
/ PRIOR FILING DATE: 2000-04-25
/ PRIOR APPLICATION NUMBER: US 09/488,725
/ PRIOR FILING DATE: 2000-01-21
/ NUMBER OF SEQ ID NOS: 63
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 51
/ LENGTH: 267
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ US-09-747-835A-51
```

```
Query Match 100.0%; Score 93; DB 10; Length 267;
Best Local Similarity 100.0%; Pred. No. 8.2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 WIDIKQFWYTPPTF 15
Db 211 WIDIKQFWYTPPTF 225
```

```
RESULT 5
US-10-012-896-879
/ Sequence 879, Application US/10012896
/ Publication No. US20020183251A1
/ GENERAL INFORMATION:
/ APPLICANT: Xu, Jiangchun
/ APPLICANT: Dillon, Davin C.
/ APPLICANT: Mitchell, Jennifer L.
/ APPLICANT: Harlocker, Susan L.
/ APPLICANT: Jiang, Yugu
/ APPLICANT: Kalos, Michael D.
```



```

; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hurai, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Mantabade, Yoshihiro
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012.896
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 879
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-012-896-879
```

```

Query Match          100.0%; Score 93; DB 9; Length 339;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WIDIKQFWMYTPPTF 15
      |||||
Db      277 WIDIKQFWMYTPPTF 291
```

```

RESULT 6
US-09-802-520-11
; Sequence 11, Application US/09802520
; Publication No. US20020187472A1
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Faris, Mary
; APPLICANT: Chen, Hwei-Mei
; APPLICANT: Ison, Craig H.
; TITLE OF INVENTION: STEAP-RELATED PROTEIN
; FILE REFERENCE: PC-0037 US
; CURRENT APPLICATION NUMBER: US/09/802.520
; CURRENT FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PERL Program
; SEQ ID NO 11
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Inctye ID No. US20020187472A1 66572948
; US-09-802-520-11
```

```

Query Match          100.0%; Score 93; DB 9; Length 339;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WIDIKQFWMYTPPTF 15
      |||||
Db      277 WIDIKQFWMYTPPTF 291
```

```

RESULT 7
US-09-895-793-879
; Sequence 879, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuguu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hurai, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895.793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 879
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-895-793-879
```

```

Query Match          100.0%; Score 93; DB 9; Length 339;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WIDIKQFWMYTPPTF 15
      |||||
Db      277 WIDIKQFWMYTPPTF 291
```

```

RESULT 8
US-09-895-814-879
; Sequence 879, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuguu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hurai, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
```

```
/ APPLICANT: Vinals de Basbols, Carlota
/ APPLICANT: Fey, Teresa
/ APPLICANT: Fanger, Gary R.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE REFERENCE: 210121.427C26
/ CURRENT APPLICATION NUMBER: US/09/895,814
/ NUMBER OF SEQ ID NOS: 990
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 879
/ LENGTH: 339
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ US-09-814-879

Query Match          100.0%; Score 93; DB 9; Length 339;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 WIDIKQFVWYTPPTF 15
Db      277 WIDIKQFVWYTPPTF 291

RESULT 9
/ US-10-011-095-2
/ Sequence 2, Application US/10011095
/ Publication No. US20030045682A1
/ GENERAL INFORMATION:
/ APPLICANT: Afar, Daniel
/ APPLICANT: Hubert, Rene S.
/ APPLICANT: Leong, Kahen
/ APPLICANT: Raitano, Arthur B.
/ APPLICANT: Saftan, Douglas C.
/ APPLICANT: Mitchell, Steve Chappell
/ TITLE OF INVENTION: ANTIBODIES IMMUNOSPECIFIC FOR STEAP1 (AS AMENDED)
/ FILE REFERENCE: 511582001610
/ CURRENT APPLICATION NUMBER: US/10/011,095
/ CURRENT FILING DATE: 2001-12-06
/ PRIOR APPLICATION NUMBER: 09/323,873
/ PRIOR FILING DATE: 1999-06-01
/ PRIOR APPLICATION NUMBER: 60/087,520
/ PRIOR FILING DATE: 1998-06-01
/ PRIOR APPLICATION NUMBER: 60/091,183
/ PRIOR FILING DATE: 1998-06-30
/ NUMBER OF SEQ ID NOS: 32
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 2
/ LENGTH: 339
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ OTHER INFORMATION: DNA
/ US-10-011-095-2

Query Match          100.0%; Score 93; DB 9; Length 339;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 WIDIKQFVWYTPPTF 15
Db      277 WIDIKQFVWYTPPTF 291

RESULT 10
/ US-10-010-667a-2
/ Sequence 2, Application US/10010667a
/ Publication No. US20030055217A1
/ GENERAL INFORMATION:
/ APPLICANT: Afar, Daniel
/ APPLICANT: Hubert, Rene S.
/ APPLICANT: Leong, Kahen
```

```
/ APPLICANT: Raitano, Arthur B.
/ APPLICANT: Saftan, Douglas C.
/ APPLICANT: Mitchell, Steve Chappell
/ TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
/ FILE REFERENCE: 511582001601
/ CURRENT APPLICATION NUMBER: US/10/010,667a
/ CURRENT FILING DATE: 2001-12-06
/ PRIOR APPLICATION NUMBER: 09/323,873
/ PRIOR FILING DATE: 1999-06-01
/ PRIOR APPLICATION NUMBER: 60/087,520
/ PRIOR FILING DATE: 1998-06-01
/ PRIOR APPLICATION NUMBER: 60/091,183
/ PRIOR FILING DATE: 1998-06-30
/ NUMBER OF SEQ ID NOS: 32
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 2
/ LENGTH: 339
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ US-10-010-667a-2

Query Match          100.0%; Score 93; DB 9; Length 339;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 WIDIKQFVWYTPPTF 15
Db      277 WIDIKQFVWYTPPTF 291

RESULT 11
/ US-09-759-143-879
/ Sequence 879, Application US/09759143
/ Patent No. US20020022248A1
/ GENERAL INFORMATION:
/ APPLICANT: Xu, Jiaqichun
/ APPLICANT: Dillon, Devin C.
/ APPLICANT: Mitcham, Jennifer L.
/ APPLICANT: Harlocker, Susan L.
/ APPLICANT: Jiang, Yuqi
/ APPLICANT: Henderson, Robert A.
/ APPLICANT: Kalos, Michael D.
/ APPLICANT: Fanger, Gary R.
/ APPLICANT: Retter, Marc W.
/ APPLICANT: Stolk, John A.
/ APPLICANT: Day, Craig H.
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Carter, Darrick
/ APPLICANT: Li, Samuel
/ APPLICANT: Wang, Aijun
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Hepler, William
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE REFERENCE: 210121.427C23
/ CURRENT APPLICATION NUMBER: US/09/759,143
/ CURRENT FILING DATE: 2001-01-12
/ NUMBER OF SEQ ID NOS: 934
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 879
/ LENGTH: 339
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ US-09-759-143-879

Query Match          100.0%; Score 93; DB 10; Length 339;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 WIDIKQFVWYTPPTF 15
Db      277 WIDIKQFVWYTPPTF 291
```

RESULT 12  
US-09-780-669-879  
; Sequence 879, Application US/09780669  
; Patent No. US20020051977A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan L.  
; APPLICANT: Jiang, Yuqin  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Stolk, John A.  
; APPLICANT: Day, Craig H.  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Carter, Darrick  
; APPLICANT: Li, Samuel  
; APPLICANT: Wang, Aijun  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Hepler, William  
; APPLICANT: Hurst, John  
; APPLICANT: McNeill, Patricia D.  
; APPLICANT: Houghton, Raymond L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.427C24  
; CURRENT APPLICATION NUMBER: US/09/780,669  
; NUMBER OF SEQ ID NOS: 943  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO: 879  
; LENGTH: 339  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-780-669-879

Query Match 100.0%; Score 93; DB 10; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WIDIKQFWYTPPTF 15  
DB 277 WIDIKQFWYTPPTF 291

RESULT 13  
US-09-822-827-879  
; Sequence 879, Application US/09822827  
; Patent No. US20020081680A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.534C1  
; CURRENT APPLICATION NUMBER: US/09/822,827  
; NUMBER OF SEQ ID NOS: 982  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO: 879  
; LENGTH: 339  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-822-827-879

Query Match 100.0%; Score 93; DB 10; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WIDIKQFWYTPPTF 15  
DB 277 WIDIKQFWYTPPTF 291

DB 277 WIDIKQFWYTPPTF 291

RESULT 14  
US-09-854-816-39  
; Sequence 39, Application US/09854816  
; Patent No. US20020151473A1  
; GENERAL INFORMATION:  
; APPLICANT: Andrew C. Braisted  
; J. Kevin Judice  
; Robert S. McDowell  
; J. Christopher Phelan  
; Melissa A. Starovaanik  
; James A. Wells  
; TITLE OF INVENTION: Constrained Helical Peptides and Methods of  
; Making Same  
; NUMBER OF SEQUENCES: 113  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/854,816  
; FILING DATE: 15-May-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/965,056  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Torchia, Phd., Timothy E.  
; REGISTRATION NUMBER: 36,700  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-8674  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 39:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 268 amino acids  
; TYPE: Amino Acid  
; TOPOLOGY: Linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 39:  
US-09-854-816-39

Query Match 48.4%; Score 45; DB 10; Length 268;  
Best Local Similarity 60.0%; Pred. No. 15;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 WIDIKQFWY 10  
DB 202 WFDITQWLMY 211

RESULT 15  
US-09-854-816-33  
; Sequence 33, Application US/09854816  
; Patent No. US20020151473A1  
; GENERAL INFORMATION:  
; APPLICANT: Andrew C. Braisted  
; J. Kevin Judice  
; Robert S. McDowell  
; J. Christopher Phelan  
; Melissa A. Starovaanik  
; James A. Wells  
; TITLE OF INVENTION: Constrained Helical Peptides and Methods of

```

;
;      NUMBER OF SEQUENCES: 113      Making Same
;      CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Genentech, Inc.
;      STREET: 1 DNA Way
;      CITY: South San Francisco
;      STATE: California
;      COUNTRY: USA
;      ZIP: 94080
;
;      COMPUTER READABLE FORM:
;      MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: Winpatin (Genentech)
;
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/09/854,816
;      FILING DATE: 15-May-2001
;      CLASSIFICATION: <Unknown>
;
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 08/965,056
;      FILING DATE: <Unknown>
;
;      ATTORNEY/AGENT INFORMATION:
;      NAME: Torchia, PhD., Timothy E.
;      REGISTRATION NUMBER: 36,700
;      REFERENCE/DOCKET NUMBER: P1005R2
;
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE: 650/225-8674
;      TELEFAX: 650/952-9881
;
;      INFORMATION FOR SEQ ID NO: 33:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 269 amino acids
;      TYPE: Amino Acid
;      TOPOLOGY: Linear
;
;      SEQUENCE DESCRIPTION: SEQ ID NO: 33:
;
US-09-854-816-33
;
Query Match      48.4%; Score 45; DB 10; Length 269;
Best Local Similarity 60.0%; Pred. No. 15;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY      1 WIDIKQFVWY 10
;
Db      203 WFDITQWLWY 212
;

```

Search completed: March 26, 2003, 17:04:17  
Job time : 11.163 secs

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: March 26, 2003, 16:44:40 ; Search time 9.59302 Seconds  
(without alignments)  
46.007 Million cell updates/sec

Title: US-10-010-667A-21  
Perfect score: 93  
Sequence: 1 WIDIKQFWYTPPTF 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	93	100.0	15	4	US-09-323-873A-21
2	93	100.0	339	4	US-09-323-873A-2
3	45	48.4	268	4	US-08-965-056-39
4	45	48.4	269	4	US-08-965-056-33
5	45	48.4	353	1	US-08-118-270-45
6	45	48.4	353	5	PCT-US93-08528-45
7	45	48.4	865	4	US-07-956-483-13
8	45	48.4	887	4	US-08-472-240A-4
9	44	47.3	403	4	US-08-887-534A-83
10	43	46.2	351	4	US-08-817-441-47
11	43	46.2	873	2	US-08-912-129A-61
12	43	46.2	877	4	US-08-817-441-102
13	42	45.2	268	4	US-08-965-056-26
14	42	45.2	268	4	US-08-965-056-68
15	42	45.2	268	4	US-08-965-056-71
16	41.5	44.6	231	4	US-08-448-489-19
17	41	44.1	237	3	US-08-388-353-642
18	41	44.1	237	3	US-08-488-551B-642
19	41	44.1	243	4	US-08-965-056-73
20	41	44.1	268	4	US-08-965-056-8
21	41	44.1	268	4	US-08-965-056-13
22	41	44.1	268	4	US-08-965-056-14
23	41	44.1	268	4	US-08-965-056-70
24	41	44.1	268	4	US-08-965-056-72
25	41	44.1	269	4	US-08-965-056-21
26	41	44.1	269	4	US-08-965-056-22
27	41	44.1	269	4	US-08-965-056-25

28	41	44.1	269	4	US-08-965-056-37	Sequence 37, Appl
29	41	44.1	270	4	US-08-965-056-67	Sequence 67, Appl
30	41	44.1	351	1	US-08-470-202-46	Sequence 46, Appl
31	41	44.1	351	1	US-08-471-770-46	Sequence 46, Appl
32	41	44.1	351	2	US-08-468-059-46	Sequence 46, Appl
33	41	44.1	351	4	US-09-109-916-46	Sequence 46, Appl
34	41	44.1	602	4	US-09-257-490-15	Sequence 15, Appl
35	41	44.1	850	2	US-08-448-603A-28	Sequence 28, Appl
36	41	44.1	850	3	US-09-134-075-28	Sequence 28, Appl
37	41	44.1	850	4	US-09-492-739-28	Sequence 28, Appl
38	41	44.1	855	1	US-08-022-835-6	Sequence 6, Appl
39	41	44.1	855	1	US-08-388-809-6	Sequence 6, Appl
40	41	44.1	855	2	US-08-647-714-6	Sequence 6, Appl
41	41	44.1	1503	4	US-08-976-255-14	Sequence 14, Appl
42	40.5	43.5	848	4	US-08-976-255-10	Sequence 10, Appl
43	40.5	43.5	1317	3	US-09-083-521-7	Sequence 7, Appl
44	40.5	43.5	1384	4	US-08-976-255-11	Sequence 11, Appl
45	40	43.0	149	4	US-09-433-428D-67	Sequence 67, Appl

## ALIGNMENTS

```

RESULT 1
US-09-323-873A-21
; Sequence 21, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahan Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; FILE REFERENCE: 129.16US02
; CURRENT APPLICATION NUMBER: US/09/323,873A
; PRIOR FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-01
; PRIOR APPLICATION NUMBER: 60/091,183
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: STEAP-1 PEPTIDE
US-09-323-873A-21

Query Match      100.0%; Score 93; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.4e+08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WIDIKQFWYTPPTF 15
DB 1 WIDIKQFWYTPPTF 15

RESULT 2
US-09-323-873A-2
; Sequence 2, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahan Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Steve Chappell Mitchell

```

TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS  
FILE REFERENCE: 129.16US2  
CURRENT APPLICATION NUMBER: US/09/323,873A  
CURRENT FILING DATE: 1999-06-01  
PRIOR APPLICATION NUMBER: 60/087,520  
PRIOR FILING DATE: 1998-06-01  
PRIOR APPLICATION NUMBER: 60/091,183  
PRIOR FILING DATE: 1998-06-30  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: PaeSeq for Windows Version 4.0  
SEQ ID NO: 2  
LENGTH: 339  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-09-323-873A-2

Query Match 100.0%; Score 93; DB 4; Length 339;  
Best Local Similarity 100.0%; Pred. No. 5,6e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WIDIKOFWYTPPTF 15  
DB 277 WIDIKOFWYTPPTF 291

RESULT 3  
US-08-965-056-39  
Sequence 39, Application US/08965056  
Patent No. 6271198  
GENERAL INFORMATION:  
APPLICANT: Andrew C. Braisted  
APPLICANT: J. Kevin Judice  
APPLICANT: Robert S. McDowell  
APPLICANT: J. Christopher Phelan  
APPLICANT: Melissa A. Starovasnik  
APPLICANT: James A. Wells  
TITLE OF INVENTION: Constrained Helical Peptides and Methods of  
TITLE OF INVENTION: Making Same  
NUMBER OF SEQUENCES: 113  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/965,056  
FILING DATE: 05-No. 6271198-1997  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, Phd., Timothy E.  
REGISTRATION NUMBER: 36,700  
REFERENCE/DOCKET NUMBER: P1005R2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/952-8674  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 268 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-08-965-056-39

Query Match 48.4%; Score 45; DB 4; Length 268;  
Best Local Similarity 60.0%; Pred. No. 11;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 WIDIKOFWY 10  
DB 202 WPDITQWLWY 211

RESULT 4  
US-08-965-056-33  
Sequence 33, Application US/08965056  
Patent No. 6271198  
GENERAL INFORMATION:  
APPLICANT: Andrew C. Braisted  
APPLICANT: J. Kevin Judice  
APPLICANT: Robert S. McDowell  
APPLICANT: J. Christopher Phelan  
APPLICANT: Melissa A. Starovasnik  
APPLICANT: James A. Wells  
TITLE OF INVENTION: Constrained Helical Peptides and Methods of  
TITLE OF INVENTION: Making Same  
NUMBER OF SEQUENCES: 113  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/965,056  
FILING DATE: 05-No. 6271198-1997  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, Phd., Timothy E.  
REGISTRATION NUMBER: 36,700  
REFERENCE/DOCKET NUMBER: P1005R2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/952-8674  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 269 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-08-965-056-33

Query Match 48.4%; Score 45; DB 4; Length 269;  
Best Local Similarity 60.0%; Pred. No. 11;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 WIDIKOFWY 10  
DB 203 WPDITQWLWY 212

RESULT 5  
US-08-118-270-45  
Sequence 45, Application US/08118270  
Patent No. 5508364  
GENERAL INFORMATION:  
APPLICANT: Murphy, Randall B.  
APPLICANT: Schuster, David I.  
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF  
NUMBER OF SEQUENCES: 348  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington

STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/118,270  
FILING DATE: 09-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/943,236  
FILING DATE: 10-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Townsend, Kevin G.  
REGISTRATION NUMBER: 34,033  
REFERENCE/DOCKET NUMBER: MURPHY=2A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 45:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 353 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-118-270-45

Query Match 48.4%; Score 45; DB 1; Length 353;  
Best Local Similarity 33.3%; Pred. No. 15;  
Matches 5; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 1 WIDIKQFWYTPPTF 15  
|::|::|::|  
Db 56 WVELYNFIMHHPWAF 70

RESULT 6  
PCT-US93-08528-45  
Sequence 45, Application PC/TUS9308528  
GENERAL INFORMATION:  
APPLICANT: New York University  
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF  
NUMBER OF SEQUENCES: 348  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/08528  
FILING DATE: 09-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/943,236  
FILING DATE: 10-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Townsend, Kevin G.  
REGISTRATION NUMBER: 34,033  
REFERENCE/DOCKET NUMBER: MURPHY=2 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528

TELEX: 248633  
INFORMATION FOR SEQ ID NO: 45:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 353 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US93-08528-45

Query Match 48.4%; Score 45; DB 5; Length 353;  
Best Local Similarity 33.3%; Pred. No. 15;  
Matches 5; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 1 WIDIKQFWYTPPTF 15  
|::|::|::|  
Db 56 WVELYNFIMHHPWAF 70

RESULT 7  
US-07-956-483-13  
Sequence 13, Application US/07956483  
Patent No. 6261799  
GENERAL INFORMATION:  
APPLICANT: KIENY, Marie-Paule  
TITLE OF INVENTION: NOVEL HYBRID, SOLUBLE AND UNCLEAVED  
TITLE OF INVENTION: gp160 VARIANT  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/956,483  
FILING DATE: 31-DEC-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO 92/19742  
FILING DATE: 12-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 91 05392  
FILING DATE: 02-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 017753-005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 865 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-07-956-483-13

Query Match 48.4%; Score 45; DB 4; Length 865;  
Best Local Similarity 60.0%; Pred. No. 36;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 WIDIKQFWY 10  
|::|::|  
Db 681 WFDITQWLWY 690

RESULT 8  
US-08-472-240A-4  
Sequence 4, Application US/08472240A  
Patent No. 6284248  
GENERAL INFORMATION:  
APPLICANT: KIENY, Marie-Paule  
TITLE OF INVENTION: NOVEL HYBRID, SOLUBLE AND UNCLEAVED  
TITLE OF INVENTION: SPI60 VARIANT  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,240A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/956,483  
FILING DATE: 31-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Teakin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 017753-055  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 887 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..854  
US-08-472-240A-4

Query Match 48.4%; Score 45; DB 4; Length 887;  
Best Local Similarity 60.0%; Pred. No. 37;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 WIDIKQFVWY 10  
Db 703 WFDITQWLMWY 712

RESULT 9  
US-08-887-534A-83  
Sequence 83, Application US/0887534A  
Patent No. 6455323  
GENERAL INFORMATION:  
APPLICANT: Holden, David W.  
TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borum  
STREET: 233 South Wacker Drive/6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/887,534A  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Rin-Laures, Li-Hsien  
REGISTRATION NUMBER: 33,547  
REFERENCE/DOCKET NUMBER: 28341/33996  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: (312) 474-6600  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 403 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-887-534A-83

Query Match 47.3%; Score 44; DB 4; Length 403;  
Best Local Similarity 58.3%; Pred. No. 24;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 DIRQFVWYTPPT 14  
Db 305 DFKVTIWTYAPT 316

RESULT 10  
US-08-817-441-47  
Sequence 47, Application US/08817441  
Patent No. 6339294  
GENERAL INFORMATION:  
APPLICANT: CHARNEAU, PIERRE  
APPLICANT: CLAVEL, FRANCOISE  
APPLICANT: BORMAN, ANDREW  
APPLICANT: QUILLENT, CAROLINE  
APPLICANT: GUETARD, DENISE  
APPLICANT: MONTAGNIER, LUC  
APPLICANT: DONOT DE SAINT-MARTIN, JACQUELINE  
APPLICANT: COHEN, JACQUES  
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF HIV-1 TYPE (OR  
TITLE OF INVENTION: SUBTYPE) ANTIGENS  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/817,441  
FILING DATE: 11-JUL-1997  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR 95/01391  
FILING DATE: 20-OCT-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: FR 9412554  
FILING DATE: 20-OCT-1994  
PRIORITY APPLICATION DATA:



APPLICATION NUMBER: FR 9502526  
FILING DATE: 03-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03260.6005-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 351 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-817-441-47

Query Match 46.2%; Score 43; DB 4; Length 351;  
Best Local Similarity 50.0%; Pred. No. 30;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WIDIKOFVWY 10  
Db 160 WIDITKWLWY 169

RESULT 11  
US-08-912-129A-61  
Sequence 61, Application US/08912129A  
Patent No. 5922533  
GENERAL INFORMATION:  
APPLICANT: VALLARI, ANADRUZELA S.  
APPLICANT: HACKETT, JOHN JR.  
APPLICANT: HICKMAN, ROBERT K.  
APPLICANT: VARITER, VINCENT A. JR.  
APPLICANT: NECKLAMS, ELIZABETH A.  
APPLICANT: GOLDEN, ALAN M.  
APPLICANT: BRENNAN, CATHERINE A.  
APPLICANT: DEYARE, SUSHIL G.  
TITLE OF INVENTION: RAPID ASSAY FOR SIMULTANEOUS DETECTION AND DIFFERENTIATIO  
NUMBER OF SEQUENCES: 89  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Abbott Laboratories  
STREET: 100 Abbott Park Road  
CITY: Abbott Park  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette, 1.44 MB  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS (Windows 95)  
SOFTWARE: Microsoft Word (ASCII format output)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/912,129A  
FILING DATE: 15-AUG-1997  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Danckers, Andreas M.  
REGISTRATION NUMBER: 32,652  
REFERENCE/DOCKET NUMBER: 6109.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847-937-9803  
TELEFAX: 847-938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 61:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 873 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
US-08-912-129A-61

Query Match 46.2%; Score 43; DB 2; Length 873;  
Best Local Similarity 50.0%; Pred. No. 75;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WIDIKOFVWY 10  
Db 682 WIDITKWLWY 691

RESULT 12  
US-08-817-441-102  
Sequence 102, Application US/08817441  
Patent No. 639294  
GENERAL INFORMATION:  
APPLICANT: CHARNEAU, PIERRE  
APPLICANT: CLAVEL, FRANCOISE  
APPLICANT: BORMAN, ANDREW  
APPLICANT: QUILBERT, CAROLINE  
APPLICANT: GUSTARD, DENISE  
APPLICANT: MONTAGNIER, LUC  
APPLICANT: COHEN, JACQUES  
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF HIV-1 TYPE (OR  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flinnegan, Henderson, Farabow, Garrett &  
STREET: Dunner, L.L.P.  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/817,441  
FILING DATE: 11-JUL-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR 95/01391  
FILING DATE: 20-OCT-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9412554  
FILING DATE: 20-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9502526  
FILING DATE: 03-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03260.6005-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 102:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 877 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-817-441-102

Query Match 46.2%; Score 43; DB 4; Length 877;  
Best Local Similarity 50.0%; Pred. No. 75;

Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 WIDIKQFWMY 10  
Db 686 WIDITKWLWY 695

## RESULT 13

US-08-965-056-26  
Sequence 26, Application US/08965056  
Patent No. 6271198  
GENERAL INFORMATION:  
APPLICANT: Andrew C. Braisted  
APPLICANT: J. Kevin Judice  
APPLICANT: Robert S. McDowell  
APPLICANT: J. Christopher Phelan  
APPLICANT: Melissa A. Starovaanik  
APPLICANT: James A. Wells  
TITLE OF INVENTION: Constrained Helical Peptides and Methods of  
TITLE OF INVENTION: Making Same  
NUMBER OF SEQUENCES: 113  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/965,056  
FILING DATE: 05-No. 6271198-1997  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, Phd., Timothy E.  
REGISTRATION NUMBER: 36,700  
REFERENCE/DOCKET NUMBER: P1005R2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/952-9881  
TELEFAX: 650/225-8674  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 268 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-08-965-056-26

Query Match 45.2%; Score 42; DB 4; Length 268;

Best Local Similarity 50.0%; Pred. No. 32;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 WIDIKQFWMY 10  
Db 202 WPDISKWLWY 211

## RESULT 14

US-08-965-056-68  
Sequence 68, Application US/08965056  
Patent No. 6271198  
GENERAL INFORMATION:  
APPLICANT: Andrew C. Braisted  
APPLICANT: J. Kevin Judice  
APPLICANT: Robert S. McDowell  
APPLICANT: J. Christopher Phelan  
APPLICANT: Melissa A. Starovaanik  
APPLICANT: James A. Wells  
TITLE OF INVENTION: Constrained Helical Peptides and Methods of  
TITLE OF INVENTION: Making Same

NUMBER OF SEQUENCES: 113  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/965,056  
FILING DATE: 05-No. 6271198-1997  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, Phd., Timothy E.  
REGISTRATION NUMBER: 36,700  
REFERENCE/DOCKET NUMBER: P1005R2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/952-9881  
TELEFAX: 650/225-8674  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 268 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-08-965-056-68

## Query Match

45.2%; Score 42; DB 4; Length 268;  
Best Local Similarity 50.0%; Pred. No. 32;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 WIDIKQFWMY 10  
Db 202 WPDISKWLWY 211

## RESULT 15

US-08-965-056-71  
Sequence 71, Application US/08965056  
Patent No. 6271198  
GENERAL INFORMATION:  
APPLICANT: Andrew C. Braisted  
APPLICANT: J. Kevin Judice  
APPLICANT: Robert S. McDowell  
APPLICANT: J. Christopher Phelan  
APPLICANT: Melissa A. Starovaanik  
APPLICANT: James A. Wells  
TITLE OF INVENTION: Constrained Helical Peptides and Methods of  
TITLE OF INVENTION: Making Same  
NUMBER OF SEQUENCES: 113  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/965,056  
FILING DATE: 05-No. 6271198-1997  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, Phd., Timothy E.  
REGISTRATION NUMBER: 36,700

REFERENCE/DOCKET NUMBER: P1005R2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 650/225-8674  
 TELEFAX: 650/952-9881  
 INFORMATION FOR SEQ ID NO: 71:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 268 amino acids  
 TYPE: Amino Acid  
 TOPOLOGY: Linear  
 US-08-965-056-71

Query Match 45.2%; Score 42; DB 4; Length 268;  
 Best Local Similarity 50.0%; Pred. No. 32;  
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 WIDIKQFWY 10  
 DB 202 WPDISRRLMY 211

Search completed: March 26, 2003, 16:51:41  
 Job time : 10.593 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 16:43:45 ; Search time 15.6977 Seconds  
(without alignments)  
196.889 Million cell updates/sec

Title: US-10-010-667A-21  
Perfect score: 93  
Sequence: 1 WIDIKQFWYTPPTF 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91	97.8	338	6 Q9GL50	Q9GL50 sus scrofa
2	80	86.0	339	11 Q9CWR7	Q9CWR7 mus musculus
3	80	86.0	339	11 Q92422	Q92422 mus musculus
4	80	86.0	339	11 Q92439	Q92439 mus musculus
5	49	52.7	448	16 Q96CK3	Q96CK3 rhizobium l
6	48	51.6	453	16 Q8UHT8	Q8UHT8 agrobacteri
7	46	49.5	122	15 Q91JL9	Q91JL9 human immun
8	46	49.5	844	15 Q97002	Q97002 human immun
9	45	48.4	107	2 Q9EVQ0	Q9EVQ0 escherichia
10	45	48.4	122	15 Q9YXR4	Q9YXR4 human immun
11	45	48.4	122	15 Q9YXR4	Q9YXR4 human immun
12	45	48.4	122	15 Q9Q1W7	Q9Q1W7 human immun
13	45	48.4	122	15 Q9Q1U8	Q9Q1U8 human immun
14	45	48.4	122	15 Q91IL6	Q91IL6 human immun
15	45	48.4	122	15 Q90DN6	Q90DN6 human immun
16	45	48.4	133	15 Q8URA0	Q8URA0 human immun

17	45	48.4	133	15 Q8UQ20	Q8UQ20 human immun
18	45	48.4	133	15 Q8UQY9	Q8UQY9 human immun
19	45	48.4	163	15 Q90E72	Q90E72 human immun
20	45	48.4	358	15 Q78120	Q78120 human immun
21	45	48.4	790	15 Q8Q2X3	Q8Q2X3 human immun
22	45	48.4	833	15 Q9QK14	Q9QK14 human immun
23	45	48.4	833	15 Q9QK13	Q9QK13 human immun
24	45	48.4	833	15 Q9QK19	Q9QK19 human immun
25	45	48.4	840	15 Q8Q2X4	Q8Q2X4 human immun
26	45	48.4	841	15 Q91V31	Q91V31 human immun
27	45	48.4	846	15 Q89292	Q89292 human immun
28	45	48.4	850	15 Q56113	Q56113 human immun
29	45	48.4	850	15 Q70003	Q70003 human immun
30	45	48.4	851	15 Q41641	Q41641 human immun
31	45	48.4	852	15 Q56567	Q56567 human immun
32	45	48.4	852	15 Q8UL55	Q8UL55 human immun
33	45	48.4	852	15 Q8UL54	Q8UL54 human immun
34	45	48.4	854	15 Q92762	Q92762 human immun
35	45	48.4	854	15 Q12008	Q12008 chimpanzee
36	45	48.4	855	15 Q902H5	Q902H5 human immun
37	45	48.4	858	15 Q87628	Q87628 chimpanzee
38	45	48.4	858	15 Q12002	Q12002 chimpanzee
39	45	48.4	859	15 Q72940	Q72940 human immun
40	45	48.4	860	15 Q9E1S7	Q9E1S7 human immun
41	45	48.4	862	15 Q12004	Q12004 chimpanzee
42	45	48.4	862	15 Q12010	Q12010 chimpanzee
43	45	48.4	863	15 Q12007	Q12007 chimpanzee
44	45	48.4	865	15 Q8Q2X5	Q8Q2X5 human immun
45	45	48.4	866	15 Q12005	Q12005 chimpanzee

#### ALIGNMENTS

RESULT 1  
Q9GL50 PRELIMINARY; PRT; 338 AA.  
ID Q9GL50  
AC Q9GL50;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
OS Six transmembrane endothelial antigen of PAEC.  
OC Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Butheria; Cetartiodactyla; Suidae; Sus.  
ON NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Nagasaka T., Boulday G., Coupel S., Coulon F., Tesson L.,  
RA Heslan J.-M., Souillou J.-P., Charreau B.;  
RT "differential gene expression in endothelial cells during TNF-alpha-  
RT and LPS-mediated activation.";  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF319659; AAG33868.1; -;  
KW Transmembrane.  
SQ SEQUENCE 338 AA; 39918 MW; ED490E86067A32B CRC64;  
Q9CWR7  
Query Match 97.8%; Score 91; DB 6; Length 338;  
Best Local Similarity 86.7%; Pred. No. 3.4e-07;  
Matches 13; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 WIDIKQFWYTPPTF 15  
Db 276 WIDIKQFWYTPPTF 290  
RESULT 2  
Q9CWR7 PRELIMINARY; PRT; 339 AA.  
AC Q9CWR7;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE 241007B19RIK protein.  
 GN STEAP OR 241007B19RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=EMBRIONIC STEM CELLS;  
 RX MEDLINE=1085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 Atakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 Blake J., Boftelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 Grotewinkel S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 Lyons P., Marchionni L., Maehima J., Mazzarelli J., Mombaerts P.,  
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 Suzuki H., Toyooka K., Wang K.H., Welz C., Whitaker C., Wilmink L.,  
 Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohnauki S.,  
 Hayashizaki Y.;  
 RA "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL; AK010437; BAB26938.1; -.  
 DR MGD; MGI:1917608; Steap.  
 SQ SEQUENCE 339 AA; 39264 MW; 3F7AB9C7520F0968 CRC64;  
 RT Query Match 86.0%; Score 80; DB 11; Length 339;  
 RT Best Local Similarity 73.3%; Pred. No. 2.4e-05;  
 RT Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 WIDIKQFWYTPPTF 15  
 Db 277 WVDVSQFWYTPPTF 291

RESULT 3  
 Q92422 PRELIMINARY; PRT; 339 AA.  
 AC Q92422; (TEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)  
 DE Dudulin.  
 GN 1010001D01RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=EMBRIONIC STEM CELLS;  
 RX MEDLINE=1085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 Atakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 Blake J., Boftelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 Grotewinkel S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 Lyons P., Marchionni L., Maehima J., Mazzarelli J., Mombaerts P.,  
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 Suzuki H., Toyooka K., Wang K.H., Welz C., Whitaker C., Wilmink L.,  
 Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohnauki S.,  
 Hayashizaki Y.;  
 RA "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL; AK010437; BAB26938.1; -.  
 DR MGD; MGI:1917608; Steap.  
 SQ SEQUENCE 339 AA; 39264 MW; 3F7AB9C7520F0968 CRC64;  
 RT Query Match 86.0%; Score 80; DB 11; Length 339;  
 RT Best Local Similarity 73.3%; Pred. No. 2.4e-05;  
 RT Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 WIDIKQFWYTPPTF 15  
 Db 277 WVDVSQFWYTPPTF 291

Query Match 86.0%; Score 80; DB 11; Length 339;  
 Best Local Similarity 73.3%; Pred. No. 2.4e-05;  
 Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 WIDIKQFWYTPPTF 15  
 Db 277 WVDVSQFWYTPPTF 291

Db 277 WVDVSQFWYTPPTF 291  
 RESULT 4  
 Q92429 PRELIMINARY; PRT; 339 AA.  
 AC Q92429; (TEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)  
 DE Six-transmembrane epithelial antigen of the prostate.  
 GN STEAP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=EMBRIONIC STEM CELLS;  
 RX MEDLINE=21371909; PubMed=11479226;  
 RA Yang D., Holt G.E., Velders M.P., Kwon E.D., Kaet W.M.;  
 RT "Murine six-transmembrane epithelial antigen of the prostate, prostate  
 stem cell antigen, and prostate-specific membrane antigen: prostate-  
 specific cell-surface and antigens highly expressed in prostate cancer of  
 transgenic adenocarcinoma mouse prostate mice."  
 RT Cancer Res. 61:5857-5860(2001).  
 RL EMBL; AF297098; AAK83126.1; -.  
 DR MGD; MGI:1917608; Steap.  
 KM Transmembrane.  
 SQ SEQUENCE 339 AA; 39318 MW; 4B26A71FF559E84F CRC64;  
 RT Query Match 86.0%; Score 80; DB 11; Length 339;  
 RT Best Local Similarity 73.3%; Pred. No. 2.4e-05;  
 RT Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 WIDIKQFWYTPPTF 15  
 Db 277 WVDVSQFWYTPPTF 291

RESULT 5  
 Q98CK3 PRELIMINARY; PRT; 448 AA.  
 AC Q98CK3; (TEMBLrel. 18, Created)  
 DT 01-OCT-2001 (TEMBLrel. 18, Last sequence update)  
 DT 01-OCT-2001 (TEMBLrel. 18, Last sequence update)  
 DT 01-OCT-2001 (TEMBLrel. 18, Last sequence update)  
 DE Alpha-glucosidase ABC transporter, substrate-binding protein.  
 GN MLL5113.  
 OS Bacterium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFP303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneke T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 Watanabe A., Ideasa K., Ishikawa A., Kawashima K., Kimura T.,  
 Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 Nishida Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
 Takeuchi C., Yamada M., Tabata S.;  
 RA "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 Mesorhizobium loti."  
 RT DNA Res. 7:331-338(2000).  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL; AF003005; BAB51618.1; -.  
 DR InterPro; IPR000567; SBP\_bac\_1.  
 DR Pfam; PF01547; SBP\_bacterial\_1; 1.  
 KM Complete proteome.  
 SQ SEQUENCE 448 AA; 48920 MW; 9B7B8569A6945D19 CRC64;  
 RT Query Match 52.7%; Score 49; DB 16; Length 448;  
 RT Best Local Similarity 53.8%; Pred. No. 4.9;  
 QY 1 WIDIKQFWYTPPTF 15  
 Db 277 WVDVSQFWYTPPTF 291

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 DIKOPWYTPPTF 15  
| | | | |  
DB 164 DVKSLVWYSPDNF 176

RESULT 6  
OSUHT8 PRELIMINARY; PRT; 453 AA.

AC 08UHT8: 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DE ABC transporter, substrate binding protein.  
GN AGLE OR ATU0591 OR AGR C.1045.  
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=176299;  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21608550; PubMed=11743193;  
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
RA Chen Y., Paulsen I.T., Bisen J.A., Karp P.D., Bovee D. Sr.,  
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,  
RA Kutyavin T., Levy R., Li M.-J., McEllland B., Palmeri A.,  
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,  
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
RA Nester E.W.;  
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
C58.";  
RL Science 294:2317-2323 (2001).  
[2]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=21608551; PubMed=11743194;  
RX Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
RA Hummel K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu F.,  
RA Wollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B.,  
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,  
RA Cielo C., Slater S.;  
RT "Genome sequence of the plant pathogen and biotechnology agent  
Agrobacterium tumefaciens C58.";  
RL Science 294:2323-2328 (2001).  
DR EMBL; AE009027; AAL41608.1; -.  
DR EMBL; AE007993; AAK86402.1; -.  
KM Complete proteome.  
SO SEQUENCE 453 AA; 49090 MW; 90896249313CD85E CRC64;

Query Match 51.6%; Score 48; DB 16; Length 453;  
Best Local Similarity 53.8%; Pred. No. 7.2;  
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 DIKOPWYTPPTF 15  
| | | | |  
DB 169 DVKSLVWYSPDNF 181

RESULT 7  
OSUHT8 PRELIMINARY; PRT; 122 AA.

AC 09JUL9: 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
GN Envelope glycoprotein (Fragment).  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroviruses; Retroviridae; Lentivirus.

OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AR3;  
RX MEDLINE=20346416; PubMed=10890362;  
RA Masciotra S., Livellara B., Belloro W., Clara L., Tanuri A., Ramos A.,  
RA Baggis J., Lal R., Pieniazek D.;  
RT "Evidence for a high frequency of HIV-1 subtype F infections among  
heterosexual population in Buenos Aires, Argentina.";  
RL AIDS Res. Hum. Retroviruses 16:1007-1014 (2000).  
DR EMBL; AF20713; AAF76832.1; -.  
DR InterPro; IPR000328; Env\_GP41.  
KW Pfam; PF00517; GP41; 1.  
FT NON TER 1 1  
FT NON TER 122 122  
SQ SEQUENCE 122 AA; 14840 MW; 10821A3BCC39F4E6 CRC64;

Query Match 49.5%; Score 46; DB 15; Length 122;  
Best Local Similarity 60.0%; Pred. No. 3.9;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 WIDIKOPWY 10  
| | | | |  
DB 112 WFDISQWLWY 121

RESULT 8  
OSUHT8 PRELIMINARY; PRT; 844 AA.

AC 097002; 01-FEB-1997 (TREMBlrel. 02, Created)  
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
GN Envelope glycoprotein (Fragment).  
RT Human immunodeficiency virus type 1.  
OC Viruses; Retroviruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=HIVBR020.17;  
RX MEDLINE=96190564; PubMed=8627686;  
RA Gao F., Morrison S.G., Robertson D.L., Thornton C.L., Craig S.,  
RA Karlsson G., Sodroski J., Morgado M., Galvao-Castro B.,  
RA von Briesen H., Beddows S., Weber J., Sharp P.M., Shaw G.M.,  
RA Hahn B.H.;  
RT "Molecular cloning and analysis of functional envelope genes from  
human immunodeficiency virus type 1 sequence subtypes A through G. The  
WHO and NIAID Networks for HIV Isolation and Characterization.";  
RL J. Virol. 70:1651-1657 (1996).  
[2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=HIVBR020.17;  
RA Allen E.E.;  
RL Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.  
DR EMBL; U27401; AAB06242.1; -.  
DR InterPro; IPR000328; Env\_GP41.  
DR InterPro; IPR000777; GP120.  
DR Pfam; PF00516; GP120; 1.  
DR Pfam; PF00517; GP41; 1.  
KM AIDS; Coat protein; Glycoprotein; Polypotein; Transmembrane.  
FT NON TER 1 1  
SQ SEQUENCE 844 AA; 95594 MW; 0E02E379CFD17691 CRC64;

Query Match 49.5%; Score 46; DB 15; Length 844;  
Best Local Similarity 60.0%; Pred. No. 30;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 WIDIKOPWY 10  
| | | | |  
DB 660 WFDISQWLWY 669

RESULT 9  
Q9EVO0 PRELIMINARY; PRT; 107 AA.  
AC Q9EVO0  
DT 01-MAR-2001 (TEMBLrel. 16, Created)  
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)  
DE 01-MAR-2001 (TEMBLrel. 16, Last annotation update)  
OS Bacteriophage 12.2 kDa protein.  
OC Bacteriophage; Proteobacteria; gamma subphylum; Enterobacteriaceae;  
OC Bacteriophage.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BJ2511;  
RX MEDLINE=20407286; PubMed=10948097;  
RA Unkempt A.; Schmidt H.;  
RT "Structural analysis of phage-borne stx genes and their flanking  
sequences in shiga toxin-producing *Escherichia coli* and *Shigella*  
dysenteriae type 1 strains."  
RL Infect. Immun. 68:4856-4864(2000).  
DR EMBL; AF251452; CAC05542.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 107 AA; 12177 MW; 8BC06C52B7D29CE6 CRC64;

Query Match 48.4%; Score 45; DB 2; Length 107;  
Best Local Similarity 46.7%; Pred. No. 5;  
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 WIDIKOFVWYTPPT 15  
DB 22 WCNKILWYQSDTF 36

RESULT 10  
Q9YXR4 PRELIMINARY; PRT; 122 AA.  
AC Q9YXR4  
DT 01-MAY-1999 (TEMBLrel. 10, Created)  
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)  
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)  
DE Envelope glycoprotein immunodominant region (Fragment).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroviridae; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RJ96BRP009;  
RA Tanuri A.; Swanson P.A.; Devare S.G.; Berro O.J.; Savedra A.;  
RA Costa L.J.; Telles J.G.; Brindeiro R.; Schable C.; Pieniazek D.;  
RA Rayfield M.;  
RT "HIV-1 subtypes among blood donors from Rio de Janeiro, Brazil."  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF034039; AAC79291.1; -.  
DR InterPro: IPR000328; Env\_GP41.  
DR Pfam: PF00517; GP41; 1.  
KW Transmembrane.  
FT NON\_TER 1  
FT NON\_TER 122  
SQ SEQUENCE 122 AA; 14739 MW; 40D4789EB8C886C2 CRC64;

Query Match 48.4%; Score 45; DB 15; Length 122;  
Best Local Similarity 60.0%; Pred. No. 5.7;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 WIDIKOFVWY 10  
DB 112 WFDITQWLMY 121

RESULT 11

Q9YX01 PRELIMINARY; PRT; 122 AA.  
AC Q9YX01  
DT 01-MAY-1999 (TEMBLrel. 10, Created)  
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)  
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)  
DE Envelope glycoprotein immunodominant region (Fragment).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroviridae; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RJ96BRP042;  
RA Tanuri A.; Swanson P.A.; Devare S.G.; Berro O.J.; Savedra A.;  
RA Costa L.J.; Telles J.G.; Brindeiro R.; Schable C.; Pieniazek D.;  
RA Rayfield M.;  
RT "HIV-1 subtypes among blood donors from Rio de Janeiro, Brazil."  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF034052; AAC79304.1; -.  
DR InterPro: IPR000328; Env\_GP41.  
DR Pfam: PF00517; GP41; 1.  
KW Transmembrane.  
FT NON\_TER 1  
FT NON\_TER 122  
SQ SEQUENCE 122 AA; 14792 MW; 7D5BB60146B8FD17 CRC64;

Query Match 48.4%; Score 45; DB 15; Length 122;  
Best Local Similarity 60.0%; Pred. No. 5.7;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 WIDIKOFVWY 10  
DB 112 WFDITQWLMY 121

RESULT 12  
Q9QIW7 PRELIMINARY; PRT; 122 AA.  
AC Q9QIW7  
DT 01-MAY-2000 (TEMBLrel. 13, Created)  
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)  
DE Envelope glycoprotein (Fragment).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroviridae; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GP41ERR04;  
RA Caride E.; Heroge K.; Larder B.; Dehertogh P.; Brindeiro R.;  
RA Machado E.; de Sa C.A.M.; Eyer W.; Passioni L.F.C.; Menezes J.A.;  
RA Calazans A.R.; Tanuri A.;  
RT "Genotyping and phenotyping analysis of B and non-B Human  
immunodeficiency virus type 1 subtypes from patients under HAART."  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF165537; AAF08482.1; -.  
DR InterPro: IPR000328; Env\_GP41.  
DR Pfam: PF00517; GP41; 1.  
KW Transmembrane.  
FT NON\_TER 1  
FT NON\_TER 122  
SQ SEQUENCE 122 AA; 14834 MW; 13711994A52A5B33 CRC64;

Query Match 48.4%; Score 45; DB 15; Length 122;  
Best Local Similarity 60.0%; Pred. No. 5.7;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 WIDIKOFVWY 10  
DB 112 WFDITQWLMY 121



RESULT 13  
ID Q90IUS PRELIMINARY; PRT; 122 AA.

AC Q90IUS; 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DE Envelope glycoprotein (Fragment).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
OX NCBI\_Taxid=11676;

RP SEQUENCE FROM N.A.  
RC STRAIN=GP4TER23;  
RA Caride E., Hertoge K., Larder B., Dehertogh P., Brindeiro R.,  
Machado E., de Sa C.A.M., Eyer W., Passioni L.F.C., Menezes J.A.,  
Calazans A.R., Tanuri A.;  
RT "Genotyping and phenotyping analysis of B and non-B Human  
immunodeficiency virus type 1 subtypes from patients under HAART";  
RL Submitted (Jul-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF165556; AAF08501.1;  
DR InterPro; IPR000328; Env\_GP41.  
KW Transmembrane.  
FT NON\_TER 1  
FT SEQUENCE 122 AA; 14733 MW; 91CD821BA7A7FECB CRC64;

Query Match 48.4%; Score 45; DB 15; Length 122;  
Best Local Similarity 60.0%; Pred. No. 5.7;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 WIDIKQFWY 10  
Db 112 WFDITQWLWY 121

RESULT 14

ID Q91LL6 PRELIMINARY; PRT; 122 AA.  
AC Q91LL6; 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DE Envelope glycoprotein (Fragment).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
OX NCBI\_Taxid=11676;

RN [1] SEQUENCE FROM N.A.  
RP STRAIN=UG1636;  
RX MEDLINE=20284721; PubMed=10826488;  
RA Downing R., Pieniazek D., Hu D.J., Biryahwaho B., Fridlund C.,  
Rayfield M.A., Sempala S.D., Lal R.B.;  
RT "Genetic characterization and phylogenetic analysis of HIV-1 subtype C  
from Uganda";  
RL AIDS Res. Hum. Retroviruses 16:815-819(2000).  
DR EMBL; AF206042; AAF81998.1;  
DR InterPro; IPR000328; Env\_GP41.  
KW Transmembrane.  
FT NON\_TER 1  
FT SEQUENCE 122 AA; 14710 MW; C1F846B32A3F4400 CRC64;

Query Match 48.4%; Score 45; DB 15; Length 122;  
Best Local Similarity 60.0%; Pred. No. 5.7;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 WIDIKQFWY 10

Db 112 WFDITQWLWY 121

RESULT 15

ID Q90DN6 PRELIMINARY; PRT; 122 AA.  
AC Q90DN6; 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DE Envelope glycoprotein (Fragment).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
OX NCBI\_Taxid=11676;

RP SEQUENCE FROM N.A.  
RC STRAIN=UG;  
RA Hu D.J., Baggs J., Downing R.G., Pieniazek D., Dorn J., Fridlund C.,  
Biryahwaho B., Sempala S.D.K., Rayfield M.A., Dondero T.J., Lal R.;  
RT "Similar distribution and continued predominance of HIV-1 subtypes A  
and D infections in Uganda";  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF307695; AAL08756.1;  
DR InterPro; IPR000328; Env\_GP41.  
KW Transmembrane.  
FT NON\_TER 1  
FT SEQUENCE 122 AA; 14728 MW; BEF43919381BDFE CRC64;

Query Match 48.4%; Score 45; DB 15; Length 122;  
Best Local Similarity 60.0%; Pred. No. 5.7;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 WIDIKQFWY 10  
Db 112 WFDITQWLWY 121

Search completed: March 26, 2003, 16:49:48  
Job time : 17.6977 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 16:44:15 ; Search time 8.54651 Seconds  
(without alignments)  
168.726 Million cell updates/sec

Title: US-10-010-667A-21

Perfect score: 93

Sequence: 1 WIDKQFWYTPPTF 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*

2: p1r2:\*

3: p1r3:\*

4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	51.6	453	2	AB97431
2	48	51.6	453	2	AB97431
3	45	48.4	31	2	A23341
4	45	48.4	358	2	S21998
5	44.5	47.8	553	2	AC0065
6	44.5	47.8	662	2	E69274
7	44	47.3	331	2	AP3544
8	44	47.3	361	2	AM1469
9	44	47.3	425	2	T27241
10	44	47.3	568	2	AB9958
11	43.5	46.8	1154	2	S69206
12	43	46.2	402	2	T14669
13	43	46.2	418	2	T14846
14	43	46.2	863	2	AS3034
15	43	46.2	877	2	S49197
16	42	45.2	133	2	G81110
17	42	45.2	321	2	E81411
18	42	45.2	396	2	S56954
19	42	45.2	462	2	B86262
20	42	45.2	639	2	E72336
21	42	45.2	685	1	EFBYS2
22	41.5	44.6	126	2	G69990
23	41.5	44.6	224	2	D69140
24	41	44.1	254	2	S21992
25	41	44.1	357	2	T44816
26	41	44.1	359	2	P84300
27	41	44.1	368	2	E90986
28	41	44.1	405	2	H85831
29	41	44.1	405	2	H85831

30	41	44.1	405	2	G64971	putative colanic a
31	41	44.1	430	2	F70016	purine permease ho
32	41	44.1	703	2	B34434	arylphorin beta ch
33	41	44.1	843	1	H44001	env polypeptide pr
34	41	44.1	847	2	T09448	env polypeptide pr
35	41	44.1	847	2	S13289	env polypeptide pr
36	41	44.1	852	2	T12016	env polypeptide pr
37	41	44.1	854	1	VCLJ51	env polypeptide pr
38	41	44.1	856	1	A44963	env polypeptide pr
39	41	44.1	929	2	G81036	isolectin-cRNA syn
40	41	44.1	929	2	B49882	probable isolectin
41	41	44.1	1042	2	S43904	hyaluronidase - Cl
42	41	44.1	1509	1	A27224	myosin heavy chain
43	40.5	43.5	313	2	A12642	flagellin [importe
44	40.5	43.5	313	2	B97425	flag protein (X964
45	40.5	43.5	1207	2	T00378	KIAA0641 protein -

#### ALIGNMENTS

##### RESULT 1

AB97431

alpha-glucosides-binding periplasmic protein aglE precursor [imported] - Agrobacterium ti

C:/Species: Agrobacterium tumefaciens

C:/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002

C:/Accession: AB97431

A:/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tume

A:/Reference number: A97359; PMID:11743194

A:/Accession: AB97431

A:/Status: preliminary

A:/Molecule type: DNA

A:/Residues: 1-453 <KUR>

A:/Cross-references: GB:AE007869; PIDN:AAK6402.1; PID:G1515534; GSPDB:GN00169

C:/Genetics:

A:/Gene: AGR C 1045

A:/Map position: circular chromosome

Query Match

Best Local Similarity 51.6%; Score 48; DB 2; Length 453;

Matches 7; Conservative 1; Mismatches 5; Gaps 0;

Db 169 DKSLVWVPEPF 181

QY 3 DKQFWYTPPTF 15

AB2649

hypothetical protein aglE [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C:/Species: Agrobacterium tumefaciens

C:/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002

C:/Accession: AB2649

A:/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:/Reference number: AB2577; PMID:11743193

A:/Status: preliminary

A:/Molecule type: DNA

A:/Residues: 1-453 <KUR>

A:/Cross-references: GB:AE008688; PIDN:AAL41608.1; PID:G17738945; GSPDB:GN00186

C:/Genetics:

A:/Gene: aglE

A:/Map position: circular chromosome

Query Match 51.6%; Score 48; DB 2; Length 453;  
 Best Local Similarity 53.8%; Pred. No. 4.1;  
 Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 3 DIKQFWYTPPTF 15  
 Db 169 DVKSLWVYDPENF 181

RESULT 3  
 A23341  
 allerGen R7 - perennial ryegrass (tentative sequence) (fragment)  
 C/Species: Lolium perenne (perennial ryegrass)  
 C/Date: 05-Jun-1987 #sequence\_revision 05-Jun-1987 #text\_change 09-Jun-2000  
 C/Accession: A23341  
 R/Cottam, G.F.; Moran, D.M.; Standing, R.  
 Biochem. J. 234, 305-310, 1986  
 A/Title: Physicochemical and immunochemical characterization of allergenic proteins from  
 A/Reference number: A23341; MUID:86242068; PMID:3718469  
 A/Accession: A23341  
 A/Molecule type: protein  
 A/Residues: 1-31 <COT>  
 C/Superfamily: expansin

Query Match 48.4%; Score 45; DB 2; Length 31;  
 Best Local Similarity 50.0%; Pred. No. 0.64;  
 Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 WIDIKQFWYTPPT 14  
 Db 18 WLDKSTWYCKPT 31

RESULT 4  
 S21998  
 envelope protein gp120/gp41 - human immunodeficiency virus type 1  
 C/Species: human immunodeficiency virus type 1, HIV-1  
 A/Variety: isolate 28  
 C/Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 01-Dec-2000  
 C/Accession: S21998; S70425  
 R/Steuler, H.; Storch-Hagenlocher, B.; Wildemann, B.; Hacke, W.  
 Submitted to the EMBL Data Library, July 1991  
 A/Description: Distinct populations of HIV-1 in blood and cerebrospinal fluid as determined  
 A/Reference number: S21990  
 A/Accession: S21998  
 A/Molecule type: DNA  
 A/Residues: 1-358 <STE2>  
 A/Cross-references: EMBL:X61359; NID:g60182; PIDN:CAA43630.1; PID:g60183  
 R/Steuler, H.; Storch-Hagenlocher, B.; Wildemann, B.  
 AIDS Res. Hum. Retroviruses 8, 53-59, 1992  
 A/Title: Distinct populations of human immunodeficiency virus type 1 in blood and cerebrospinal  
 A/Reference number: S70417; MUID:92144209; PMID:1736940  
 A/Accession: S70425  
 A/Molecule type: preliminary  
 A/Status: preliminary  
 A/Residues: 1-222 'X', 224-358 <STE2>  
 A/Cross-references: EMBL:X61359; NID:g60182; PIDN:CAA43630.1; PID:g60183  
 C/Superfamily: type B retrovirus env polyprotein

Query Match 48.4%; Score 45; DB 2; Length 358;  
 Best Local Similarity 60.0%; Pred. No. 9.8;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 WIDIKQFWY 10  
 Db 174 WPDITQWLMY 183

RESULT 5  
 AC0065  
 conserved hypothetical protein YPO0524 (imported) - Yersinia pestis (strain CO92)  
 C/Species: Yersinia pestis

C/Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
 C/Accession: AC0065  
 R/Parhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.;  
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, I.  
 Nature 413, 523-527, 2001  
 A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A/Reference number: AB0001; MUID:21470413; PMID:11586360  
 A/Accession: AC0065  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-553 <KUR>  
 A/Cross-references: GB:AL590842; PIDN:CAC89382.1; PID:g15978619; GSPDB:GN00175  
 C/Genetics:  
 A/Gene: YPO0524

Query Match 47.8%; Score 44.5; DB 2; Length 553;  
 Best Local Similarity 53.8%; Pred. No. 19;  
 Matches 7; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

Qy 1 WIDIKQFWYTPPT 13  
 Db 540 WDFPKS-AMFTTP 551

RESULT 6  
 E69274  
 acetyl-CoA synthetase (acc-1) homolog - Archaeoglobus fulgidus  
 C/Species: Archaeoglobus fulgidus  
 C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 21-Jan-2000  
 C/Accession: E69274  
 R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson,  
 .; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;  
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A/Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.  
 Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
 A/Reference number: A69250; MUID:98049343; PMID:9389475  
 A/Accession: E69274  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-662 <LTB>  
 A/Cross-references: GB:AE001092; GB:AE00782; NID:g2689415; PIDN:AAB91033.1; PID:g2650444  
 C/Superfamily: acetate-CoA ligase; acetate-CoA ligase homology  
 F:145-621/Domain: acetate-CoA ligase homology <ACT>

Query Match 47.8%; Score 44.5; DB 2; Length 662;  
 Best Local Similarity 41.2%; Pred. No. 24;  
 Matches 7; Conservative 6; Mismatches 1; Indels 3; Gaps 1;

Qy 2 IDIKQFWYTPPT 15  
 Db 307 WDFKGDRLWYSPSM 323

RESULT 7  
 AF3544  
 hypothetical membrane spanning protein BMEI10279 (imported) - Brucella melitensis (strain  
 C/Species: Brucella melitensis  
 C/Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
 C/Accession: AF3544  
 R/DelVecchio, V.G.; Kaparat, V.; Redkar, R.J.; Patra, G.; Muier, C.; Los, T.; Ivanova, I.  
 .; Mazur, M.; Goldman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessc  
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A/Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
 A/Reference number: AD3252; PMID:11756688  
 A/Accession: AF3544  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-331 <KUR>  
 A/Cross-references: GB:AE008918; PIDN:AAL53521.1; PID:g17984427; GSPDB:GN00191  
 A/Experimental source: strain 16M

C:Genetics:  
A:Gene: BMEI10279  
A:Map position: 11

Query Match 47.3%; Score 44; DB 2; Length 331;  
Best Local Similarity 38.5%; Pred. No. 13;  
Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 3 DIKQFVWYTPPT 15  
DB 117 NVKDFVWYTPPT 129

# RESULT 8

Internal protein homology lin0295 [imported] - *Listeria innocua* (strain C1p11262)

C:Species: *Listeria innocua*  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

C:Accession: AH1469

R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Jones, L.M.; Kerst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkut, G.; Madueno, E.; Maitournam, A.; Ma

Ok, C.; Schueller, T.; Simoes, N.; Tiller, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlund, A.

A:Title: Comparative genomics of *Listeria* species.

A:Reference number: AB1077; PMID:11679669

A:Accession: AH1469

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-361 <GLA>

A:Cross-references: GB:AL592022; PIDN:G16412724; GSPDB:GN00178

A:Experimental source: strain C1p11262

C:Genetics:

A:Gene: lin0295

Query Match 47.3%; Score 44; DB 2; Length 361;  
Best Local Similarity 42.9%; Pred. No. 14;  
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 IDIKQFVWYTPPT 15  
DB 310 VTLHDVWYTPPT 323

# RESULT 9

T27241  
Hypothetical protein Y57G11C.31 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T27241

R:McMurray, A.

submitted to the EMBL Data Library, September 1997

A:Reference number: Z20330

A:Accession: T27241

A:Status: preliminary; translated from GB/EMBL/DDDB

A:Molecule type: DNA

A:Residues: 1-425 <HIL>

A:Cross-references: EMBL:Z29291; PIDN:CAH16530.1; GSPDB:GN00022; CESP:Y57G11C.31

A:Experimental source: clone Y57G11C

C:Genetics:

A:Gene: CESP:Y57G11C.31

A:Map position: 4

A:introns: 160/3; 187/2; 309/2; 361/3

Query Match 47.3%; Score 44; DB 2; Length 425;  
Best Local Similarity 46.7%; Pred. No. 17;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 WIDIKQFVWYTPPT 15  
DB 146 WPEIQDFHWPPT 160

RESULT 10  
AB8958  
acetyl-CoA synthetase [imported] - *Staphylococcus aureus* (strain N315)

C:Species: *Staphylococcus aureus*

C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 17-May-2002

C:Accession: AB8958

R:Kuroda, M.; Ono, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; I

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramoto, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.

A:Reference number: AB8758; PMID:11418146

A:Accession: AB8958

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-568 <KUR>

A:Cross-references: GB:BA000018; PID:G13701528; PIDN:BA82822.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: acsA

Query Match 47.3%; Score 44; DB 2; Length 568;  
Best Local Similarity 58.3%; Pred. No. 24;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 DIKQFVWYTPPT 14  
DB 292 DFKVWYTPPT 303

# RESULT 11

S69206  
regulator protein white collar 1 - *Neurospora crassa*

C:Species: *Neurospora crassa*

C:Date: 21-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 11-Jan-2002

C:Accession: S69206

R:Ballario, P.; Victorioso, P.; Magrelli, A.; Talora, C.; Cabibbo, A.; Macino, G.

EMBO J. 15, 1650-1657, 1996

A:Title: White collar-1, a central regulator of blue light responses in *Neurospora*, is a

A:Reference number: S69206; PMID:8612589

A:Accession: S69206

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1154 <BAL>

A:Cross-references: EMBL:X94300; NID:G1279576; PID:G1480115

C:Genetics:

A:introns: 967/3

C:Superfamily: GATA-type zinc finger homology

C:Keywords: zinc finger

P:932-991/Domain: GATA-type zinc finger homology <GZF>

Query Match 46.8%; Score 43.5; DB 2; Length 1154;  
Best Local Similarity 66.7%; Pred. No. 64;  
Matches 8; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY 3 DIKQFVWYTPPT 14  
DB 529 DIQYTWTPPT 539

# RESULT 12

T14669  
P-loop protein - *Yersinia pestis* plasmid pMT1

C:Species: *Yersinia pestis*

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T14669

R:Hu, P.; Elliott, J.; McCready, P.; Skowronski, E.; Garnea, J.; Kobayashi, A.; Carrano, submitted to the EMBL Data Library, March 1998

A:Description: Structural organization of virulence determinants in three *Yersinia pestis*

A:Reference number: Z18168

A:Accession: T14669

A/Status: preliminary; translated from GB/EMBL/DBD  
 A/Molecule type: DNA  
 A/Residues: 1-402 <HUP>  
 C/Cross-references: EMBL:AF053947; NID:G2996286; PID:G2996304; PIDN:AAC13184.1  
 C/Genetics: plasmid pMT1  
 A/Genome: plasmid pMT1

Query Match 46.2%; Score 43; DB 2; Length 402;  
 Best Local Similarity 63.6%; Pred. No. 24;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 KQFVWTPPTF 15  
 |||||  
 Db 38 KORWVYVAPY 48

## RESULT 13

T14946  
 hypothetical protein Y1030 - Yersinia pestis plasmid pMT1

C/Species: Yersinia pestis  
 C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 22-Oct-1999  
 C/Accession: T14946

R/Indier: L.E.; Plano, G.V.; Burland, V.; Mayhew, G.F.; Blattner, F.R.  
 Infect. Immun. 66, 5731-5742, 1998

A/Title: Complete DNA sequence and detailed analysis of the Yersinia pestis KIMS plasmid  
 A/Reference number: Z18268; MUID:99043898; PMID:9826348

A/Accession: T14946

A/Status: preliminary; translated from GB/EMBL/DBD

A/Molecule type: DNA

A/Residues: 1-418 <LIN>

A/Cross-references: EMBL:AF074611; NID:G3883003; PID:G3883031; PIDN:AAC82691.1  
 C/Genetics:  
 A/Gene: Y1030  
 A/Genome: plasmid pMT1

Query Match 46.2%; Score 43; DB 2; Length 418;  
 Best Local Similarity 63.6%; Pred. No. 25;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 KQFVWTPPTF 15  
 |||||  
 Db 54 KORWVYVAPY 64

## RESULT 14

A53034  
 gag polyprotein - human immunodeficiency virus type 1 (strain Ant70)

C/Species: human immunodeficiency virus type 1, HIV-1  
 C/Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 07-May-1999  
 C/Accession: A53034

R/Vanden Haesevelde, M.; Decourt, J.L.; De Leye, R.J.; Vanderborght, B.; van der Groen,  
 J. Virol. 68, 1586-1596, 1994

A/Title: Genomic cloning and complete sequence analysis of a highly divergent African hu

A/Reference number: A53034; MUID:94149849; PMID:8107220

A/Accession: A53034

A/Status: preliminary

A/Molecule type: genomic RNA

A/Residues: 1-863 <VAN>

A/Cross-references: GB:I02587

C/Superfamily: type B retrovirus env polyprotein

C/Keywords: polyprotein

## RESULT 15

S49197

envelope protein precursor - human immunodeficiency virus type 1 (fragment)  
 C/Species: human immunodeficiency virus type 1, HIV-1  
 C/Date: 19-Mar-1997 #sequence\_revision 24-Jul-1997 #text\_change 26-Aug-1999  
 C/Accession: S49197

R/Chamneau, P.; Borman, A.M.; Oullent, C.; Guetard, D.; Charnaret, S.; Cohen, J.; Remy, C

submitted to the EMBL Data Library, July 1994

A/Description: Isolation and envelope sequence of a highly divergent HIV-1 isolate: défir

A/Reference number: S49197

A/Accession: S49197

A/Molecule type: DNA

A/Residues: 1-877 <CHA>

A/Cross-references: EMBL:X80020; NID:G510516; PIDN:CA56323.1; PID:G510517

A/Experimental source: Isolate VAN

C/Superfamily: type B retrovirus env polyprotein

C/Keywords: glycoprotein; capsid protein; coat protein; transmembrane protein

F:1-30/Domain: signal sequence #status predicted <Sig>

F:31-535/Product: coat protein gp120 #status predicted <CP1>

F:536-877/Product: coat protein gp41 #status predicted <CP2>

F:898-716/Domain: transmembrane #status predicted <TMN>

F:59,88,139,148,159,184,188,198,230,235,242,263,270,277,292,302,333,345,357,367,396,404,4

Query Match 46.2%; Score 43; DB 2; Length 877;  
 Best Local Similarity 50.0%; Pred. No. 57;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 WIDIKQFVWY 10  
 |||||  
 Db 686 WLDITKWLWY 695

Search completed: March 26, 2003, 16:50:42  
 Job time: 10.5465 secs

GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 16:38:14 ; Search time 29.1279 Seconds  
(without alignments)  
68.620 Million cell updates/sec

Title: US-10-010-667a-21  
Perfect score: 93  
Sequence: 1 WIDIKQFWYTPPTF 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: A\_Geneseq\_101002.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	93	100.0	15	AA158200	Human STRAP-1 pept
2	93	100.0	15	AAE02788	Extracellular loop
3	93	100.0	339	AA158194	Human STRAP-1 prot
4	93	100.0	339	AAU69927	Human prostate CDN
5	93	100.0	339	AAU78845	Human protein SEQ
6	93	100.0	339	AAU01282	P788P amino acid s
7	93	100.0	339	ABG61813	Prostate cancer-as
8	93	100.0	339	ABB95387	Human P789P protei
9	93	100.0	374	AAU79829	Human protein SEQ
10	93	100.0	375	AAE02780	Human six transmem

11	81	87.1	339	20	AAH86309	Kidney injury asso
12	47	50.5	51	22	AAU21212	Human novel foetal
13	45	48.4	268	19	AAV22843	SEQ ID NO. 39 from
14	45	48.4	268	23	ABG68314	Envelope protein g
15	45	48.4	269	19	AAU22837	SEQ ID NO. 33 from
16	45	48.4	269	23	ABG68308	Envelope protein g
17	45	48.4	353	15	AAU48724	G-protein coupled
18	45	48.4	353	17	AAU02696	G-protein coupled
19	45	48.4	846	21	AAH69345	HIV-1 non-subtype
20	45	48.4	861	19	AAU43074	HIV-1 gp120 protei
21	45	48.4	866	9	AAH80966	HIV protein HT7.
22	44	47.3	100	22	AAU04068	Human polypeptide
23	44	47.3	403	20	AAU97719	Staphylococcus aur
24	44	47.3	568	22	AAU33996	Staphylococcus aur
25	44	47.3	568	22	AAU37036	Staphylococcus aur
26	43	46.2	62	22	AAU43980	Propionibacterium
27	43	46.2	119	23	AAU09876	Novel human secret
28	43	46.2	147	23	AAU79366	Zcy87. Homo sapi
29	43	46.2	148	22	AAU67508	Amino acid sequenc
30	43	46.2	159	22	AAU67507	Amino acid sequenc
31	43	46.2	163	19	AAU69321	Anti-HIV-1 group O
32	43	46.2	171	21	AAU77261	Protein VAV relati
33	43	46.2	173	21	AAU77259	Protein AN770 rela
34	43	46.2	290	22	AAU54642	Propionibacterium
35	43	46.2	474	21	AAU77371	HIV-1 group O env
36	43	46.2	641	22	ABH61933	Drosophila melanog
37	43	46.2	873	20	AAU09501	HIV-1 Group O env
38	43	46.2	873	20	AAU06985	Amino acid sequenc
39	43	46.2	873	20	AAU77376	HIV-1 group O isol
40	43	46.2	877	17	AAU07205	HIV-1 group O beta
41	42	45.2	268	19	AAU22872	SEQ ID NO. 68 from
42	42	45.2	268	19	AAU22875	SEQ ID NO. 71 from
43	42	45.2	268	19	AAU22830	SEQ ID NO. 26 from
44	42	45.2	268	23	ABG68301	Envelope protein g
45	42	45.2	268	23	ABG68343	Envelope protein g

#### ALIGNMENTS

RESULT 1	AA158200	AA158200 standard; peptide; 15 AA.
ID	AA158200	AA158200 standard; peptide; 15 AA.
AC	AA158200	
XX	14-MAR-2000	(first entry)
XX	Human STRAP-1 peptide, corresponding to STRAP-1 extracellular region 3.	
XX	Serpentine transmembrane antigen of the prostate; STRAP-1; prostate;	
KW	transmembrane domain; type IIIa membrane protein; expression; cancer;	
KW	prostate cancer; bladder cancer; colon cancer; pancreatic cancer;	
KW	ovarian cancer; tumour antigen; immunisation; immune response;	
KW	cellular; humoral; anticancer vaccine; antibody; detection; diagnosis;	
KW	prognosis; monitoring; susceptibility; therapeutic inhibitor;	
KW	drug targeting; recombinant protein.	
OS	Synthetic.	
OS	Homo sapiens.	
XX	MO9962941-AA2.	
XX	09-DEC-1999.	
PD	01-JUN-1999;	99MO-US12157.
XX	01-JUN-1999;	99MO-US12157.
XX	01-JUN-1998;	98US-0087520.
PR	30-JUN-1998;	98US-0091183.
XX	(UROC-) UROGENESYS INC.	
PA	(AFAR/) AFAR D E.	
PA	(HUBE/) HUBERT R S.	

PA (LEON/) LEONG K.  
 PA (RAIT/) RAITANO A. B.  
 PA (SAFE/) SAFEFRAN D. C.  
 XX  
 PI Afar DE, Hubert RS, Leong K, Raitano AB, Saffran DC.  
 XX  
 DR WPI; 2000-072832/06.  
 XX  
 PT Novel proteins useful as diagnostic markers and therapeutic targets,  
 XX particularly for prostatic cancer -  
 XX  
 PS Disclosure; Page 22; 83pp; English.  
 XX  
 CC Sequences AAY58198-Y58200 represent synthetic peptides that correspond  
 CC to the extracellular regions of STRAP-1 (serpentine transmembrane  
 CC antigen of the prostate; AAY58194). These peptides were used to raise  
 CC monoclonal anti-STRAP-1 antibodies. STRAP-1 is the prototype  
 CC member of the STRAP family of proteins (AAY58194-Y58197) which  
 CC exhibit a high degree of structural conservation, but which show  
 CC no significant structural homology to known human proteins. The STRAP-1  
 CC gene has been localised to chromosome 7p22. STRAP-1 is thought to be a  
 CC type IIa membrane protein and is expressed predominantly in prostate  
 CC cells in normal human tissues. Structurally, STRAP-1 is a 339 amino  
 CC acid protein characterised by six transmembrane domains and  
 CC intracellular N- and C-termini, suggesting that it folds in a  
 CC "serpentine" manner into three extracellular and two intracellular loops.  
 CC STRAP-1 mRNA and protein expression is maintained at high levels and  
 CC throughout all stages of prostate cancer. STRAP-1 mRNA and/or protein is  
 CC also overexpressed in certain other cancers, including bladder, colon,  
 CC pancreatic and ovarian cancer. The function of the STRAP proteins is not  
 CC known. They may be ion channels (from the presence of six transmembrane  
 CC domains, a feature which is shared by certain ion channels) or  
 CC gap-junction proteins (from immunohistochemical staining). STRAP-1 and  
 CC STRAP-2 are cell-surface tumour antigens. Immunisation with a STRAP  
 CC protein induces cellular and humoral immune responses against  
 CC STRAP-expressing cells. STRAP proteins may be used to identify  
 CC specific-binding agents, to produce anticancer vaccines and to generate  
 CC specific antibodies. The antibodies may be used for detection, prognosis,  
 CC and monitoring of cancers (or susceptibility to cancer), as therapeutic  
 CC inhibitors or to target therapeutic agents to their site of action. STRAP  
 CC nucleic acids may be used for recombinant protein production, as  
 CC diagnostic and prognostic reagents, for identifying STRAP-expressing  
 CC cells for screening inhibitors of STRAP expression and for therapeutic  
 CC modulation/inhibition of STRAP expression. Since high levels of STRAP  
 CC proteins are exposed on the cell surface, they are easily targeted by  
 CC systemically administered agents, and because they are expressed mainly  
 CC on prostatic epithelial cells, agents targeted to them should have  
 CC minimal side effects on other tissues.  
 CC  
 XX  
 SQ Sequence 15 AA;  
 QY  
 DB 1 WIDIKQFWVWTPPTF 15  
 1 WIDIKQFWVWTPPTF 15  
 RESULT 2  
 ID AAE02788  
 AA02788 standard; peptide; 15 AA.  
 XX  
 AC AAE02788,  
 XX  
 DT 06-AUG-2001 (first entry)  
 XX  
 DE Extracellular loop #3 of human STRAP-1, suitable for cloning into pFC.  
 XX  
 KW Human; cytostatic; antiproliferative; vaccine; gene therapy;  
 KW six transmembrane epithelial antigen of the prostate-1; STRAP-1;  
 KW chromosome 7p22.3; cancer; prostate; colon; bladder; pancreatic;

KW ovarian; lung; extracellular loop; serpentine transmembrane antigen.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200140276-A2.  
 XX  
 PD 07-JUN-2001.  
 XX  
 PF 06-DEC-2000; 2000WO-US33040.  
 XX  
 PR 06-DEC-1999; 99US-0455486.  
 XX  
 PA (UROG-) UROGENESIS INC.  
 XX  
 PI Afar DEH, Hubert RS, Raitano AB, Saffran DC, Mitchell SC, Faris M;  
 XX Jakobovits A;  
 XX  
 DR WPI; 2001-367804/38.  
 XX  
 PT New STRAP (six transmembrane epithelial antigen of the prostate)  
 PT proteins, expressed in human cancers, useful for detecting and treating  
 PT cancer -  
 PS  
 XX Example 19; Page 102; 187pp; English.  
 XX  
 CC The present invention relates to human six transmembrane epithelial  
 CC antigen of the prostate (STRAP) protein. STRAP is a member of cell  
 CC surface serpentine transmembrane antigens. STRAP gene is used in gene  
 CC therapy. Inhibiting the development or progression of a cancer (eg.  
 CC prostate, colon, bladder, lung, ovarian and pancreatic) expressing STRAP  
 CC or inhibiting growth or killing cells expressing STRAP in a patient;  
 CC comprises administering a vaccine composition to the patient. Treating  
 CC a patient with a cancer that expresses STRAP, or inhibiting growth or  
 CC killing cells expressing STRAP, comprises administering to the patient a  
 CC vector encoding single chain monoclonal antibody that comprises the  
 CC variable domains of the heavy and light chains of the monoclonal antibody  
 CC that specifically binds to STRAP, such that the vector delivers the  
 CC single chain monoclonal antibody coding sequence to the cancer cells and  
 CC the encoded single chain monoclonal antibody is expressed  
 CC intracellularly. The present sequence is extracellular loop of STRAP-1  
 CC suitable for cloning into pFC, which is used in the invention. STRAP-1  
 CC gene is located on chromosome 7p22.3.  
 CC  
 XX  
 SQ Sequence 15 AA;  
 QY  
 DB 1 WIDIKQFWVWTPPTF 15  
 1 WIDIKQFWVWTPPTF 15  
 RESULT 3  
 ID AAY58194  
 AAY58194 standard; Protein; 339 AA.  
 XX  
 AC AAY58194;  
 XX  
 DT 14-MAR-2000 (first entry)  
 XX  
 DE Human STRAP-1 protein.  
 XX  
 KW Serpentine transmembrane antigen of the prostate; STRAP-1; prostate;  
 KW transmembrane domain; type IIa membrane protein; expression; cancer;  
 KW prostate cancer; bladder cancer; colon cancer; pancreatic cancer;  
 KW ovarian cancer; tumour antigen; immunisation; immune response;  
 KW cellular; humoral; anticancer vaccine; antibody; detection; diagnosis;  
 KW prognosis; monitoring; susceptibility; therapeutic inhibitor;  
 KW drug targeting; recombinant protein.  
 XX  
 OS Homo sapiens.



Key	Location/Qualifiers
Region	1..69
Region	/note= "Intracellular region 1"
Domain	70..91
Region	/note= "Transmembrane domain 1"
Domain	92..113
Region	/note= "Extracellular region 1 (AAV58198)"
Domain	114..136
Region	/note= "Transmembrane domain 2"
Domain	137..162
Region	/note= "Intracellular region 2"
Domain	163..184
Region	/note= "Transmembrane domain 3"
Domain	185..218
Region	/note= "Extracellular region 2 (AAV58199)"
Domain	219..241
Region	/note= "Transmembrane domain 4"
Domain	242..253
Region	/note= "Intracellular region 3"
Domain	252..276
Region	/note= "Transmembrane domain 5"
Domain	277..291
Region	/note= "Extracellular region 3 (AAV58200)"
Domain	292..313
Region	/note= "Transmembrane domain 6"
Domain	314..339
Region	/note= "Intracellular region 4"
Region	WO9962941-A2.
Region	09-DEC-1999.
Region	01-JUN-1999;
Region	99WO-US12157.
Region	01-JUN-1998;
Region	98US-0087520.
Region	30-JUN-1998;
Region	98US-0091183.
Region	(UROG-) UROGENESYS INC.
Region	(AFAR/) AFAR D E.
Region	(HUBE/) HUBERT R. S.
Region	(LEON/) LEONG K. S.
Region	(RAIT/) RAITANO A. B.
Region	(SAFF/) SAFFRAN D C.
Region	Afar DE, Hubert RS, Leong K, Raitano AB, Saffran DC;
Region	WPI, 2000-072832/06.
Region	N-PSDB; AA249395, AA249396.
Region	Novel proteins useful as diagnostic markers and therapeutic targets, particularly for prostatic cancer -
Region	Claim 1; Fig 1A; 83pp; English.
Region	This sequence represents a novel human protein, STRAP-1 (serpentine transmembrane antigen of the prostate). STRAP-1 is the prototype member of the STRAP family of proteins (AAV58194-Y58197) which exhibit a high degree of structural conservation, but which show no significant structural homology to known human proteins. The STRAP-1 gene has been localized to chromosome 7p22. STRAP-1 is thought to be a type IIIa membrane protein and is expressed predominantly in prostate cells in normal human tissues. Structurally, STRAP-1 is a 339 amino acid protein characterised by six transmembrane domains and intracellular N- and C-termini, suggesting that it folds in a "serpentine" manner into three extracellular and two intracellular loops throughout all stages of prostate cancer. STRAP-1 mRNA and/or protein is also overexpressed in certain other cancers, including bladder, colon, pancreatic and ovarian cancer. The function of the STRAP proteins is not known. They may be ion channels (from the presence of six transmembrane domains), a feature which is shared by certain ion channels or gap-junction proteins (from immunohistochemical staining). STRAP-1 and

CC		STRAP-2 are cell-surface tumour antigens. Immunisation with a STRAP
CC		protein induces cellular and humoral immune responses against
CC		STRAP-expressing cells. STRAP proteins may be used to identify
CC		specific-binding agents, to produce anticancer vaccines and to generate
CC		specific antibodies. The antibodies may be used for detection, prognosis,
CC		and monitoring of cancers (or susceptibility to cancer), as therapeutic
CC		inhibitors or to target therapeutic agents to their site of action. STRAP
CC		nucleic acids may be used for recombinant protein production, as
CC		diagnostic and prognostic reagents, for identifying STRAP-expressing
CC		cells for screening inhibitors of STRAP expression and for therapeutic
CC		modulation/inhibition of STRAP expression. Since high levels of STRAP
CC		proteins are exposed on the cell surface, they are easily targeted by
CC		systemically administered agents, and because they are expressed mainly
CC		on prostatic epithelial cells, agents targeted to them should have
CC		minimal side effects on other tissues.
SQ	Sequence	339 AA;
OY		1 WIDIKQFVWTPPTPF 15              Db     277 WIDIKQFVWTPPTP 291
RESULT 4		
AU69927	ID AU69927 standard; Protein:	339 AA.
AC	AAU69927;	
XX	30-JAN-2002 (first entry)	
DZ		
DE	Human prostate cDNA encoded protein #72.	
KM	Human; prostate cancer; cytostatic; immunosupplant; tumour; immunogen.	
XN	Homo sapiens.	
OS	WO200173032-A2.	
PX	04-OCT-2001.	
PD	27-MAR-2001; 2001WO-US09919.	
PF		
XX	27-MAR-2000; 2000US-0536857.	
PR	09-MAY-2000; 2000US-0568100.	
PR	12-MAY-2000; 2000US-0570737.	
PR	13-JUN-2000; 2000US-0593793.	
PR	27-JUN-2000; 2000US-0605793.	
PR	10-AUG-2000; 2000US-0636215.	
PR	29-AUG-2000; 2000US-0651236.	
PR	06-SEP-2000; 2000US-0657279.	
PR	02-OCT-2000; 2000US-0679426.	
PR	10-OCT-2000; 2000US-0685156.	
PA	(CORI-) CORIXA CORP.	
PI	Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD; Pi Fanger GR, Reiter MW, Stolk JA, Day CH, Vedvik TS, Carter D; Pi Li SX, Wang A, Skeiky VAW, Hepler WT, Henderson RA;  DR WPI, 2001-639232/73. DX N-PADB; AAS64160.	
PT	New human prostate-specific polypeptides and poly nucleotides useful for	
PS	the diagnosis and treatment of cancer, especially prostate cancer -	
XX	Claim 2; Page 549; 579pp; English.	
CC	The invention relates to isolated prostate-specific	

CC polynucleotides, polypeptides, fusion proteins of the polypeptides,  
CC antibodies raised against the polypeptides (or antigenic epitopes  
CC derived from them) and antigen-presenting cells expressing the  
CC polypeptides. The antibodies are useful for detecting the presence of  
CC cancer, especially prostate cancer. The polypeptides, polynucleotides and  
CC the antigen-presenting cells are useful for stimulating and/or expanding  
CC T cells specific for a tumour protein, and for inhibiting the development  
CC of cancer especially prostate cancer. Compositions comprising the  
CC polynucleotide and/or polypeptide are useful for stimulating an immune  
CC response, and for treating cancer. The oligonucleotide is useful for  
CC detecting cancer. The present sequence is a prostate specific  
CC polypeptide of the invention.

SQ Sequence 339 AA;

Query Match 100.0%; Score 93; DB 22; Length 339;  
Best Local Similarity 100.0%; Pred. No. 8.4e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WIDIKQFVWYTPPTF 15  
Db 277 WIDIKQFVWYTPPTF 291

RESULT 5  
AAM78845  
ID AAM78845 standard; Protein; 339 AA.

XX AAM78845;

XX 06-NOV-2001 (first entry)

XX Human protein SEQ ID NO 1507.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
XX tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorder; arthritis; inflammation.

XX Homo sapiens.

XX WO200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US04098.

XX 03-FEB-2000; 2000US-0496914.

XX 27-APR-2000; 2000US-0560875.

XX 20-JUN-2000; 2000US-0598075.

XX 19-JUL-2000; 2000US-0620325.

XX 01-SEP-2000; 2000US-0654936.

XX 15-SEP-2000; 2000US-0663561.

XX 20-OCT-2000; 2000US-0693325.

XX 30-NOV-2000; 2000US-0728422.

XX (HYSR-) HYSRQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;

XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

XX Xue AJ, Yang J, Wejhrman T, Goodrich R;

XX MPI; 2001-476283/51.

XX N-PSDB; AAKS1978.

XX DR

XX XX

CC The invention relates to polynucleotides (AAKS1456-AAKS3435) and the  
CC encoded polypeptides (AAM78333-AAW80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce

CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAKS2581), 2111 (AAKS2582) and 3666  
CC (AAW80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.

SQ Sequence 339 AA;

Query Match 100.0%; Score 93; DB 22; Length 339;  
Best Local Similarity 100.0%; Pred. No. 8.4e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WIDIKQFVWYTPPTF 15  
Db 277 WIDIKQFVWYTPPTF 291

RESULT 6  
AAM01282  
ID AAM01282 standard; Protein; 339 AA.

XX AAM01282;

XX 04-OCT-2001 (first entry)

XX P789P amino acid sequence.

XX Human; prostate cancer; prostate-specific; diagnosis; vaccine;  
XX cytostatic; gene therapy; metastasis.

XX Homo sapiens.

XX WO200151633-A2.

XX 19-JUL-2001.

XX 16-JAN-2001; 2001WO-US01574.

XX 14-JAN-2000; 2000US-0483672.

XX (CORI-) CORIXA CORP.

XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;

XX Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;

XX Wang A, Meagher MJ;

XX MPI; 2001-425873/45.

XX DR

XX XX

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

CC Claim 2; Page 510-512; 543pp; English.  
CC The present invention describes polynucleotide sequences (I) which encode  
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,  
CC and can be used in vaccine production and gene therapy. (I), (II),  
CC antibodies to (II), fusion proteins comprising (II), and isolated  
CC T cells prepared using (I) or (II) are used treat cancer in a patient.  
CC (I) and the antibodies are also used in the detection of cancer in a  
CC patient. The cancer that is diagnosed or treated is particularly  
CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or  
CC (I) can be used for monitoring the progression of cancer in a patient.  
CC (I) and (II) can also be used to improve diagnostic and therapeutic  
CC methods for prostate cancer. They can indicate the level of metastasis  
CC as well as the prostate volume. AAK93357 to AAK93944 and AAM01115 to  
CC AAM01318 represent polynucleotide and amino acid sequences used in the

CC exemplification of the present invention.  
XX  
SQ Sequence 339 AA;  
Query Match 100.0%; Score 93; DB 22; Length 339;  
Best Local Similarity 100.0%; Pred. No. 8.4e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MIDIKOFWYTPPTF 15  
DB 277 MIDIKOFWYTPPTF 291  
RESULT 7  
ABG61813  
ID ABG61813 standard; Proteins; 339 AA.  
XX  
AC ABG61813;  
XX  
DT 15-AUG-2002 (first entry)  
XX  
DE Prostate cancer-associated protein #14.  
XX  
KW Prostate cancer; prostate tumour tissue; human; mammal; cytostatic.  
XX  
OS Mammalia.  
XX  
PN WO200230268-A2.  
XX  
PD 18-APR-2002.  
XX  
PF 12-OCT-2001; 2001WO-US32045.  
XX  
PR 13-OCT-2000; 2000US-0687576.  
XX  
PR 08-DEC-2000; 2000US-0733288.  
XX  
PR 08-DEC-2000; 2000US-0733742.  
XX  
PR 24-JAN-2001; 2001US-263957P.  
XX  
PR 16-MAR-2001; 2001US-276791P.  
XX  
PR 16-MAR-2001; 2001US-276888P.  
XX  
PR 06-APR-2001; 2001US-281922P.  
XX  
PR 24-APR-2001; 2001US-286214P.  
XX  
PR 30-APR-2001; 2001US-0847046.  
XX  
PR 04-MAY-2001; 2001US-288589P.  
XX  
PA (EOSB-) EOS BIOTECHNOLOGY INC.  
XX  
PI Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;  
XX  
XX  
DR WPI; 2002-471335/50.  
XX  
DR N-PSDB; ABK92128.  
XX  
PT Detecting a prostate cancer-associated transcript in a cell in a  
XX  
PT patient; useful for diagnosing prostate cancer (PC) or screening  
XX  
PT modulators of PC, by determining if prostate cancer-associated genes  
XX  
PT are expressed in a prostate tissue -  
XX  
PS Claim 27; Page 312; 436pp; English.  
XX  
XX  
CC The present invention relates to methods of detecting a prostate  
XX  
CC cancer-associated transcript in a cell from a patient. The method  
XX  
CC comprises contacting a biological sample from the patient with  
XX  
CC prostate cancer-associated polynucleotides (designated PC genes) that  
XX  
CC selectively hybridize to a sequence that is at least 80% identical  
XX  
CC to them. The prostate cancer-associated polynucleotide sequences  
XX  
CC are differentially expressed in prostate tumour tissue or in  
XX  
CC prostate cancer and are derived from the tissues of various  
XX  
CC organisms such as humans or other mammals (e.g. mice, sheep and dogs).  
XX  
CC The methods of the invention are useful for diagnosing and treating  
XX  
CC prostate cancer in mammals. The prostate cancer-associated genes are  
XX  
CC useful for diagnosing or treating prostate cancer, as well as for  
XX  
CC identifying modulators of prostate cancer or agents that inhibit  
XX  
CC prostate cancer. The nucleic acid sequences are particularly useful  
XX  
CC in gene therapy, as a vaccine or in antisense applications.

CC ABG61800-ABG61944 represent prostate cancer-associated proteins.  
XX  
SQ Sequence 339 AA;  
Query Match 100.0%; Score 93; DB 23; Length 339;  
Best Local Similarity 100.0%; Pred. No. 8.4e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MIDIKOFWYTPPTF 15  
DB 277 MIDIKOFWYTPPTF 291  
RESULT 8  
ABG61800  
ID ABG61800 standard; Proteins; 339 AA.  
XX  
AC ABG61800;  
XX  
DT 19-JUL-2002 (first entry)  
XX  
DE Human P789P protein SEQ ID NO 879.  
XX  
KW Human; cancer; prostate cancer; vaccine; cytostatic; immunostimulant;  
XX  
KW gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN US2002022248-A1.  
XX  
PD 21-FEB-2002.  
XX  
PF 12-JAN-2001; 2001US-0759143.  
XX  
PR 25-FEB-1997; 97US-0806099.  
XX  
PR 01-AUG-1997; 97US-0904804.  
XX  
PR 09-FEB-1998; 98US-0020956.  
XX  
PR 25-FEB-1998; 98US-0030607.  
XX  
PR 14-JUL-1998; 98US-0115453.  
XX  
PR 23-SEP-1998; 98US-0159812.  
XX  
PR 15-JUN-1999; 99US-0232149.  
XX  
PR 09-APR-1999; 99US-0288946.  
XX  
PR 13-JUL-1999; 99US-0352616.  
XX  
PR 12-NOV-1999; 99US-0439313.  
XX  
PR 18-NOV-1999; 99US-0443686.  
XX  
PR 14-JAN-2000; 2000US-0483672.  
XX  
PR 27-MAR-2000; 2000US-0536857.  
XX  
PR 09-MAY-2000; 2000US-0568100.  
XX  
PR 12-MAY-2000; 2000US-0570737.  
XX  
PR 13-JUN-2000; 2000US-0593793.  
XX  
PR 27-JUN-2000; 2000US-0605783.  
XX  
PR 10-AUG-2000; 2000US-0636215.  
XX  
PR 29-AUG-2000; 2000US-0651236.  
XX  
PR 06-SEP-2000; 2000US-0657279.  
XX  
PR 02-OCT-2000; 2000US-0679426.  
XX  
PR 10-OCT-2000; 2000US-0685166.  
XX  
XX  
PA (XUJ/) XU J.  
XX  
PA (DIL/) DILON D C.  
XX  
PA (MIT/) MITCHAM J L.  
XX  
PA (HAR/) HARLOCKER S L.  
XX  
PA (JIAN/) JIANG Y.  
XX  
PA (KALO/) KALOS M D.  
XX  
PA (FANG/) FANGER G R.  
XX  
PA (RETT/) RETTER M W.  
XX  
PA (STOL/) STOLK J A.  
XX  
PA (DAY/) DAY C H.  
XX  
PA (VEDV/) VEDVICK T S.  
XX  
PA (CART/) CARTER D.  
XX  
PA (LISX/) LI S X.  
XX  
PA (WANG/) WANG A.  
XX  
PA (SKEI/) SKEIKY Y A W.  
XX  
PA (HEPL/) HEPLER W T.

PA (HEND/) HENDERSON R A.  
XX  
XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;  
PI Panzer GR, Rafter MW, Stolk JA, Day CH, Vedvick TS, Carter D;  
PI Li SX, Wang A, Skelky YAW, Hepler WT, Henderson RA;  
XX WPI, 2002-255649/30.  
XX  
XX New prostate-specific polynucleotides for diagnosing and treating  
PT diseases, in particular prostate cancer, and as markers for the  
PT progression of cancer -  
XX  
XX Claim 2: SEQ ID NO 879; 87bp; English.  
XX  
XX The present invention provides prostate-specific coding sequences and  
CC their encoded proteins. These can be used in the diagnosis and treatment  
CC of cancers, particularly prostate cancer. The present sequence is a  
CC protein described in the invention.  
XX  
XX  
SQ Sequence 339 AA;  
Query Match 100.0%; Score 93; DB 23; Length 339;  
Best Local Similarity 100.0%; Pred. No. 8.4e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 WIDIKQFWWTPTPTF 15  
DB 277 WIDIKQFWWTPTPTF 291  
RESULT 9  
AAM79829  
ID AAM79829 standard; Protein; 374 AA.  
XX  
XX AAM79829;  
AC  
XX  
XX 06-NOV-2001 (first entry)  
DT  
XX  
XX Human protein SEQ ID NO 3475.  
DE  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200157190-A2.  
PN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 05-FEB-2001; 2001MO-US04098.  
PF  
XX  
XX 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Tang YT, Liu C, Dymnac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;  
XX  
XX WPI; 2001-476283/51.  
DR N-PSDB; AAK52962.  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -

XX Claim 20; Page 350; 6221pp; English.  
PS  
XX  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52561), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
XX  
XX  
SQ Sequence 374 AA;  
Query Match 100.0%; Score 93; DB 22; Length 374;  
Best Local Similarity 100.0%; Pred. No. 9.2e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 WIDIKQFWWTPTPTF 15  
DB 312 WIDIKQFWWTPTPTF 326  
RESULT 10  
AAE02780  
ID AAE02780 standard; Protein; 375 AA.  
XX  
XX AAE02780;  
AC  
XX  
XX 06-AUG-2001 (first entry)  
DT  
XX  
XX Human six transmembrane epithelial antigen of prostate (STREP)-1 protein.  
DE  
XX  
XX Human; cytostatic; antiproliferative; vaccine; gene therapy;  
KW six transmembrane epithelial antigen of the prostate-1; STREP-1;  
KW chromosome 7p22.3; cancer; prostate; colon; bladder; lung; ovarian;  
KW pancreatic.  
XX  
XX Homo sapiens.  
OS  
XX  
XX  
FH Key Location/Qualifiers  
FT Region 14..28  
FT /label= Immunogenic\_peptide #1  
FT /label= Transmembrane\_domain #1  
FT /label= Transmembrane\_domain #1  
FT Region 86..94  
FT /label= HLA-A2\_binding\_peptide #2  
FT /label= Transmembrane\_domain #2  
FT Region 114..136  
FT /label= Transmembrane\_domain #2  
FT /label= Transmembrane\_domain #5  
FT Region 158..166  
FT /label= HLA-A2\_binding\_peptide #5  
FT /label= Transmembrane\_domain #5  
FT Region 163..184  
FT /label= Transmembrane\_domain #3  
FT /label= Transmembrane\_domain #3  
FT Region 165..173  
FT /label= HLA-A2\_binding\_peptide #1  
FT /label= Transmembrane\_domain #1  
FT Domain 219..241  
FT /label= Transmembrane\_domain #4  
FT Domain 254..276  
FT /label= Transmembrane\_domain #5  
FT Region 262..270  
FT /label= HLA-A2\_binding\_peptide #3  
FT Domain 292..313  
FT /label= Transmembrane\_domain #6  
FT Region 302..310  
FT /label= HLA-A2\_binding\_peptide #4  
FT Misc-difference 339..340  
FT /note= "Encoded by TTGTACAAAT"

```

XX XX WO200140276-A2.
XX XX
XX XX 07-JUN-2001.
XX XX
XX XX 06-DEC-2000; 2000WO-US33040.
XX XX
XX XX 06-DEC-1999; 99US-0455486.
XX XX
XX XX (UROG-) UROGENESYS INC.
XX XX
XX XX Afar DEH, Hubert RS, Raitano AB, Saffran DC, Mitchell SC, Faris M,
XX XX Jakobovits A;
XX XX WPI; 2001-367804/38.
XX XX N-PSDB; AAD07067.
XX XX
XX XX New STEAP (six transmembrane epithelial antigen of the prostate)
XX XX PT proteins, expressed in human cancers, useful for detecting and treating
XX XX PT cancer -
XX XX
XX XX Example 2; Fig 1A-1B; 187pp; English.
XX XX
XX XX The present sequence is human six transmembrane epithelial antigen of
XX XX CC the prostate (STEAP)-1 protein of clone 10. STEAP-1 is a member of cell
XX XX CC surface serpentine transmembrane antigens. STEAP-1 gene is located on
XX XX CC chromosome 7p22.3 and is used in gene therapy. Inhibiting the development
XX XX CC or progression of a cancer (eg, prostate, colon, bladder, lung, ovarian
XX XX CC and pancreatic) expressing STEAP or inhibiting growth or killing cells
XX XX CC expressing STEAP in a patient, comprises administering a vaccine
XX XX CC composition to the patient. Treating a patient with a cancer that
XX XX CC expresses STEAP, or inhibiting growth or killing cells expressing STEAP,
XX XX CC comprises administering to the patient a vector encoding single chain
XX XX CC monoclonal antibody that comprises the variable domains of the heavy and
XX XX CC light chains of the monoclonal antibody that specifically binds to STEAP,
XX XX CC such that the vector delivers the single chain monoclonal antibody coding
XX XX CC sequence to the cancer cells and the encoded single chain monoclonal
XX XX CC antibody is expressed intracellularly.
XX XX CC Note: The present sequence is also shown in sequence listing of the
XX XX CC specification, but it lacks amino acid residues at its N-terminal end.
XX XX
XX XX Sequence 375 AA;
XX XX
XX XX Query Match 100.0%; Score 93; DB 22; Length 375;
XX XX Best Local Similarity 100.0%; Pred. No. 9.2e-06;
XX XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX XX
XX XX QY 1 WIDIKQFWWTPTTF 15
XX XX Db 277 WIDIKQFWWTPTTF 291
XX XX
XX XX RESULT 11
XX XX AAW86309
XX XX ID AAW86309 standard; Protein; 339 AA.
XX XX
XX XX AC AAW86309;
XX XX
XX XX 01-MAR-1999 (first entry)
XX XX
XX XX Kidney injury associated molecule HM018 protein.
XX XX
XX XX Kidney injury associated molecule; kidney injury related molecule;
XX XX KM KIM; tissue growth promotion; regeneration; renal condition;
XX XX KM acute renal failure; acute nephritis; tumour.
XX XX
XX XX Ratus sp.
XX XX
XX XX PN MO9853071-A1.
XX XX
XX XX PD 26-NOV-1998.
XX XX
XX XX 22-MAY-1998; 98WO-US10547.

```

```

XX XX 23-MAY-1997; 97US-0047491.
XX XX PR 23-MAY-1997; 97US-0047490.
XX XX
XX XX (BIOJ ) BIOGEN INC.
XX XX
XX XX Cate RL, Hession CA, Sanicola-Nadel M, Wei H,
XX XX PI
XX XX WPI; 1999-045312/04.
XX XX DR N-PSDB; AAW80586.
XX XX
XX XX Kidney injury-associated molecule, KIM, polypeptides - upregulated
XX XX PT in injured or regenerating tissues, useful to promote tissue growth
XX XX PT and regeneration, especially to treat renal conditions
XX XX
XX XX Claim 17; Page 57-58; 213pp; English.
XX XX
XX XX The present sequence represents a kidney injury associated molecule
XX XX CC (KIM) protein. KIM proteins can be administered therapeutically
XX XX CC by expressing KIM encoding polynucleotides, to promote growth and/or
XX XX CC survival of damaged tissue (e.g. renal tissue), since the KIM proteins
XX XX CC are upregulated in injured or regenerating (especially renal) tissues.
XX XX CC KIM fusion proteins, conjugates, antibodies and vectors can also be used
XX XX CC therapeutically, e.g. these or the KIM proteins may be included with an
XX XX CC acceptable carrier in pharmaceutical compositions, useful for therapy/
XX XX CC prophylaxis of conditions associated with dysfunction/dysregulation of
XX XX CC KIM genes or proteins, especially renal diseases or impairments of renal
XX XX CC function in humans (e.g. acute renal failure, acute nephritis). The
XX XX CC polynucleotides can be used to produce antisense sequences which, when
XX XX CC internalised into cells, can disrupt expression of a cellular KIM gene,
XX XX CC also useful in therapy (e.g. to block the growth of tumours dependent on
XX XX CC KIM for growth) or compositions. The proteins and polynucleotides are
XX XX CC useful diagnostically e.g. to detect and quantify renal injury/disease
XX XX CC (indicative of increased risk, or presence of, renal injury or impaired
XX XX CC function), or abnormal responses to tissue injury (indicative of
XX XX CC increased risk, or presence of, an autoimmune response or abnormal
XX XX CC tissue growth arising from/affecting renal tissue). The proteins can
XX XX CC also be used to locate KIM-producing cells (especially specific loci,
XX XX CC e.g. tissue masses abnormally producing/expressing KIM such as tumours
XX XX CC arising from/affecting renal tissue), by contacting cells with an
XX XX CC imageable KIM-binding reagent and imaging reagent accumulation.
XX XX
XX XX Sequence 339 AA;
XX XX
XX XX Query Match 87.1%; Score 81; DB 20; Length 339;
XX XX Best Local Similarity 80.0%; Pred. No. 0.00047;
XX XX Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX XX
XX XX QY 1 WIDIKQFWWTPTTF 15
XX XX Db 277 WVIDISQFWWTPTTF 291
XX XX
XX XX RESULT 12
XX XX AAU21212
XX XX ID AAU21212 standard; Protein; 51 AA.
XX XX
XX XX AC AAU21212;
XX XX
XX XX 17-DEC-2001 (first entry)
XX XX
XX XX Human novel foetal antigen, SEQ ID NO 1456.
XX XX
XX XX Human; foetal tissue antigen; antiinflammatory; neuroprotective;
XX XX KM immunomodulator; cardiovascular; cytostatic; nephrochotropic;
XX XX KM cardiovascular; autoimmune disease; rheumatoid arthritis;
XX XX KM hyperproliferative disorder; breast neoplasm; cancer;
XX XX KM cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
XX XX KM cerebral ischaemia; angiogenesis; nervous system disorder;
XX XX KM Alzheimer's disease; infection; ocular disorder; corneal infection;
XX XX KM wound healing; epithelial cell proliferation; food additive.
XX XX
XX XX Homo sapiens.
XX XX
XX XX

```

XX W0200155312-A2.  
XX 02-AUG-2001.  
PF 17-JAN-2001, 2001WO-US01321.  
XX  
PR 31-JAN-2000, 2000US-0179065.  
PR 04-FEB-2000, 2000US-0180628.  
PR 24-FEB-2000, 2000US-0184664.  
PR 02-MAR-2000, 2000US-0186350.  
PR 16-MAR-2000, 2000US-0189874.  
PR 17-MAR-2000, 2000US-0190076.  
PR 18-APR-2000, 2000US-0198123.  
PR 19-MAY-2000, 2000US-0205515.  
PR 07-JUN-2000, 2000US-0209467.  
PR 28-JUN-2000, 2000US-0214886.  
PR 30-JUN-2000, 2000US-0215135.  
PR 07-JUL-2000, 2000US-0216647.  
PR 07-JUL-2000, 2000US-0216880.  
PR 11-JUL-2000, 2000US-0217487.  
PR 11-JUL-2000, 2000US-0217496.  
PR 14-JUL-2000, 2000US-0218290.  
PR 26-JUL-2000, 2000US-0220963.  
PR 26-JUL-2000, 2000US-0220964.  
PR 14-AUG-2000, 2000US-0224518.  
PR 14-AUG-2000, 2000US-0224519.  
PR 14-AUG-2000, 2000US-0225213.  
PR 14-AUG-2000, 2000US-0225214.  
PR 14-AUG-2000, 2000US-0225266.  
PR 14-AUG-2000, 2000US-0225267.  
PR 14-AUG-2000, 2000US-0225268.  
PR 14-AUG-2000, 2000US-0225270.  
PR 14-AUG-2000, 2000US-0225447.  
PR 14-AUG-2000, 2000US-0225757.  
PR 14-AUG-2000, 2000US-0225758.  
PR 14-AUG-2000, 2000US-0225759.  
PR 18-AUG-2000, 2000US-0226279.  
PR 22-AUG-2000, 2000US-0226681.  
PR 22-AUG-2000, 2000US-0226868.  
PR 22-AUG-2000, 2000US-0227182.  
PR 23-AUG-2000, 2000US-0227009.  
PR 30-AUG-2000, 2000US-0228924.  
PR 01-SEP-2000, 2000US-0229287.  
PR 01-SEP-2000, 2000US-0229343.  
PR 01-SEP-2000, 2000US-0229344.  
PR 01-SEP-2000, 2000US-0229345.  
PR 05-SEP-2000, 2000US-0229509.  
PR 05-SEP-2000, 2000US-0229513.  
PR 06-SEP-2000, 2000US-0230437.  
PR 06-SEP-2000, 2000US-0230438.  
PR 08-SEP-2000, 2000US-0231242.  
PR 08-SEP-2000, 2000US-0231243.  
PR 08-SEP-2000, 2000US-0231244.  
PR 08-SEP-2000, 2000US-0231413.  
PR 08-SEP-2000, 2000US-0231414.  
PR 08-SEP-2000, 2000US-0232080.  
PR 08-SEP-2000, 2000US-0232081.  
PR 12-SEP-2000, 2000US-0231968.  
PR 14-SEP-2000, 2000US-0232397.  
PR 14-SEP-2000, 2000US-0232398.  
PR 14-SEP-2000, 2000US-0232399.  
PR 14-SEP-2000, 2000US-0232460.  
PR 14-SEP-2000, 2000US-0232401.  
PR 14-SEP-2000, 2000US-0233053.  
PR 14-SEP-2000, 2000US-0233064.  
PR 14-SEP-2000, 2000US-0233065.  
PR 21-SEP-2000, 2000US-0234223.  
PR 21-SEP-2000, 2000US-0234274.  
PR 25-SEP-2000, 2000US-0234997.  
PR 25-SEP-2000, 2000US-0234998.  
PR 26-SEP-2000, 2000US-0235484.  
PR 27-SEP-2000, 2000US-0235834.

PR 27-SEP-2000, 2000US-0235836.  
PR 29-SEP-2000, 2000US-0236327.  
PR 29-SEP-2000, 2000US-0236367.  
PR 29-SEP-2000, 2000US-0236368.  
PR 29-SEP-2000, 2000US-0236369.  
PR 29-SEP-2000, 2000US-0236370.  
PR 02-OCT-2000, 2000US-0236802.  
PR 02-OCT-2000, 2000US-0237037.  
PR 02-OCT-2000, 2000US-0237038.  
PR 02-OCT-2000, 2000US-0237039.  
PR 02-OCT-2000, 2000US-0237040.  
PR 13-OCT-2000, 2000US-0239935.  
PR 13-OCT-2000, 2000US-0239937.  
PR 20-OCT-2000, 2000US-0240960.  
PR 20-OCT-2000, 2000US-0241221.  
PR 20-OCT-2000, 2000US-0241785.  
PR 20-OCT-2000, 2000US-0241786.  
PR 20-OCT-2000, 2000US-0241787.  
PR 20-OCT-2000, 2000US-0241808.  
PR 20-OCT-2000, 2000US-0241809.  
PR 20-OCT-2000, 2000US-0241826.  
PR 01-NOV-2000, 2000US-0244617.  
PR 08-NOV-2000, 2000US-0246474.  
PR 08-NOV-2000, 2000US-0246475.  
PR 08-NOV-2000, 2000US-0246476.  
PR 08-NOV-2000, 2000US-0246477.  
PR 08-NOV-2000, 2000US-0246478.  
PR 08-NOV-2000, 2000US-0246523.  
PR 08-NOV-2000, 2000US-0246524.  
PR 08-NOV-2000, 2000US-0246525.  
PR 08-NOV-2000, 2000US-0246526.  
PR 08-NOV-2000, 2000US-0246527.  
PR 08-NOV-2000, 2000US-0246528.  
PR 08-NOV-2000, 2000US-0246532.  
PR 08-NOV-2000, 2000US-0246610.  
PR 08-NOV-2000, 2000US-0246611.  
PR 08-NOV-2000, 2000US-0246613.  
PR 17-NOV-2000, 2000US-0249207.  
PR 17-NOV-2000, 2000US-0249208.  
PR 17-NOV-2000, 2000US-0249209.  
PR 17-NOV-2000, 2000US-0249210.  
PR 17-NOV-2000, 2000US-0249211.  
PR 17-NOV-2000, 2000US-0249212.  
PR 17-NOV-2000, 2000US-0249214.  
PR 17-NOV-2000, 2000US-0249215.  
PR 17-NOV-2000, 2000US-0249216.  
PR 17-NOV-2000, 2000US-0249217.  
PR 17-NOV-2000, 2000US-0249218.  
PR 17-NOV-2000, 2000US-0249244.  
PR 17-NOV-2000, 2000US-0249245.  
PR 17-NOV-2000, 2000US-0249246.  
PR 17-NOV-2000, 2000US-0249265.  
PR 17-NOV-2000, 2000US-0249297.  
PR 17-NOV-2000, 2000US-0249299.  
PR 17-NOV-2000, 2000US-0249300.  
PR 01-DEC-2000, 2000US-0250160.  
PR 01-DEC-2000, 2000US-0250391.  
PR 05-DEC-2000, 2000US-0251030.  
PR 05-DEC-2000, 2000US-0251988.  
PR 05-DEC-2000, 2000US-0251989.  
PR 06-DEC-2000, 2000US-0251479.  
PR 08-DEC-2000, 2000US-0251856.  
PR 08-DEC-2000, 2000US-0251858.  
PR 08-DEC-2000, 2000US-0251869.  
PR 08-DEC-2000, 2000US-0251989.  
PR 08-DEC-2000, 2000US-0251990.  
PR 11-DEC-2000, 2000US-0254097.  
PR 05-JAN-2001, 2001US-0259678.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX

PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-48782/53.  
DR N-PSDB; AAS34032.  
XX  
PT New polynucleotides and polypeptides for diagnosing, treating,  
PT preventing or diagnosing e.g. diseases or disorders of the nervous,  
PT musculoskeletal, excretory, gastrointestinal, reproductive, and  
PT respiratory systems  
XX  
PS Claim 11; SEQ ID No 1456; 642pp; English.  
XX  
XX The invention relates to novel nucleic acids encoding novel human foetal  
CC antigens. The nucleic acids and proteins are used to prevent, treat (e.g.  
CC by gene therapy) or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. The antibodies to the antigens can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. Numerous  
CC examples of diseases and disorders treated by the nucleic acids and  
CC proteins are given in the specification. The present sequence

Query Match 50.5%; Score 47; DB 22; Length 51;  
Best Local Similarity 46.7%; Pred. No. 7.2;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 WIDIKQFWYPTPT 15  
DB 30 WPDLOQSFYCPPLSF 44

RESULT 13  
AAV22843  
ID AAV22843 standard; Protein; 268 AA.  
XX  
AC AAV22843;  
XX  
DT 19-AUG-1999 (first entry)  
XX  
DE SEQ ID NO. 39 from WO9820036.  
XX  
XX HIV: gp41 protein; constrained helical peptide; HIV infection;  
KW vaccine; antibody; viral membrane fusion; viral infectivity;  
KW ligand affinity purification; protein A replacement;  
KW immunoglobulin purification; epitope mimic.  
XX  
OS Human immunodeficiency virus.  
XX  
PN WO9820036-A1.  
XX  
PD 14-MAY-1998.  
XX  
PF 05-NOV-1997; 97WO-US20069.  
XX  
PR 16-JUN-1997; 97US-0876698.  
PR 06-NOV-1996; 96US-0743698.  
XX  
PA (GETH ) GENENTECH INC.

XX Braisted A, Judice JK, McDowell RS, Phelan JC, Starovasnik MA;  
PI Wells JA;  
XX WPI; 1998-286866/25.  
XX  
XX Production of constrained helical peptide(s) by linking side chains  
PT on termini of octapeptide - derived from human immunodeficiency  
PT virus gp41 protein, useful in vaccines for treatment and prevention  
PT of infection  
XX  
PS Claim 11; Page 180-181; 279pp; English.  
XX  
XX Peptides AAV22805-Y22917 are derived from Human immunodeficiency virus  
CC (HIV). Specifically, AAV22810-Y22910 are derived from gp41 proteins  
CC of known HIV virus strains (AAV22810, AAV22871, AAV22880, AAV22888 and  
CC AAV22903 represent consensus sequences of various sections of the gp41  
CC protein). Sequences derived from the peptides are used to produce  
CC constrained helical peptides of the invention. The constrained helical  
CC peptide is produced by synthesizing an octapeptide in which both terminal  
CC amino acids have a side-chain that includes a group able to form an amide  
CC bond, and cyclising the octapeptide by reacting the specified side-chain  
CC residues with a difunctional linker to produce two amide bonds.  
CC The constrained helical peptides are used to treat or prevent HIV  
CC infection, especially as vaccines that generate antibodies that  
CC prevent viral membrane fusion or infectivity. Vaccines may contain  
CC constrained helical peptides derived from several different strains of  
CC HIV. The antibodies are also useful for diagnosing HIV infection. Other  
CC uses for the constrained helical peptides are in affinity purification  
CC of ligands (particularly where complete binding protein is not readily  
CC available), e.g. replacements for protein A in immunoglobulin  
CC purification; as epitope mimics for antibody production; for isolation  
CC of synthetic antibody clones from phage display libraries, or as stable  
CC forms of "floppy" peptides or proteins.  
XX  
SQ Sequence 268 AA;

Query Match 48.4%; Score 45; DB 19; Length 268;  
Best Local Similarity 60.0%; Pred. No. 69;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 WIDIKQFWY 10  
DB 202 WPDITQWLWY 211

RESULT 14  
ABG68314  
ID ABG68314 standard; Protein; 268 AA.  
XX  
AC ABG68314;  
XX  
DT 07-OCT-2002 (first entry)  
XX  
DE Envelope protein gp41 from HIV clade B strain #33.  
XX  
XX HIV: glycoprotein; gp41; antigen; helical conformation;  
KW virus-induced membrane fusion; acquired immunodeficiency syndrome; AIDS;  
KW viral envelope protein; vaccine; virucide; anti-HIV.  
XX  
OS Human immunodeficiency virus type 1 clade B.  
XX  
PN US6271198-B1.  
XX  
PD 07-AUG-2001.  
XX  
PF 05-NOV-1997; 97US-0965056.  
XX  
PR 16-JUN-1997; 97US-049787P.  
PR 06-NOV-1996; 96US-0743698.  
PR 16-JUN-1997; 97US-0876698.  
XX  
PA (GETH ) GENENTECH INC.

XX Braisted AC, Judice JK, McDowell RS, Pheelan JC, Starovashnik MA;  
 PI Wells JA;  
 XX  
 DR WPI; 2002-487624/52.  
 XX  
 PT New cyclic peptides from human immune deficiency virus gp41, useful for  
 PT treatment or prevention of HIV infection, are constrained to have  
 PT alpha-helical conformation -  
 XX  
 PS Disclosure; Column 189-192; 175pp; English.  
 XX  
 CC The invention relates to cyclic peptides (A) with a constrained helical  
 CC conformation, derived from gp41 (glycoprotein 41), a viral envelope  
 CC protein of human immunodeficiency virus (HIV). The cyclic  
 CC peptides have formulas given in the specification part of which are  
 CC derived from a consensus sequence of gp41 derived from HIV clades A, B,  
 CC C, D, E or O. The peptides are used to cause induction of a specific  
 CC immune response, resulting in antibodies that prevent virus-induced  
 CC membrane fusion. The peptides are used to treat subjects with, or at risk  
 CC of, HIV infection, either as antifection/anti-infection agents or,  
 CC preferably where associated with a carrier, as an immunogen (including as  
 CC vaccine) to raise antibodies. The antibodies may be used for diagnosis or  
 CC prevention/treatment of HIV infection (i.e. acquired immunodeficiency  
 CC syndrome, AIDS), e.g. prevention of mother-to-child transmission or in  
 CC cases of health care accidents. The peptides can be based on specific HIV  
 CC strains, e.g. breakthrough isolates of HIV that have developed during  
 CC vaccine trials, so a combination of them should cover a wide range of  
 CC protection. The present sequence is gp41 protein from a particular  
 CC HIV clade used to derive a consensus sequence of gp41.  
 CC  
 SQ Sequence 268 AA;  
 QY 1 WIDIKQFVWY 10  
 DB 202 WFDITQWLWY 211  
 Query Match 48.4%; Score 45; DB 23; Length 268;  
 Best Local Similarity 60.0%; Pred. No. 69;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 RESULT 15  
 AAY22837  
 ID AAY22837 standard; Protein; 269 AA.  
 AC AAY22837;  
 XX  
 DT 19-AUG-1999 (first entry)  
 XX  
 DE SEQ ID NO. 33 from WO9820036.  
 XX  
 KW HIV; gp41 protein; constrained helical peptide; HIV infection;  
 KW vaccine; antibody; viral membrane fusion; viral infectivity;  
 KW ligand affinity purification; protein A replacement;  
 KW immunoglobulin purification; epitope mimic.  
 XX  
 OS Human immunodeficiency virus.  
 XX  
 PN WO9820036-A1.  
 PD 14-MAY-1998.  
 XX  
 PF 05-NOV-1997; 97WO-US20069.  
 XX  
 PR 16-JUN-1997; 97US-0876698.  
 PR 06-NOV-1996; 96US-0743698.  
 XX  
 PA (GETH ) GENENTECH INC.  
 PI Braisted A, Judice JK, McDowell RS, Pheelan JC, Starovashnik MA;  
 PI Wells JA;  
 XX

DR WPI; 1998-286866/25.  
 XX  
 PT Production of constrained helical peptide(s) by linking side chains  
 PT on termini of octa-peptide - derived from human immunodeficiency  
 PT virus gp41 protein, useful in vaccines for treatment and prevention  
 PT of infection  
 XX  
 PS Claim 11; Page 174-175; 279pp; English.  
 XX  
 CC Peptides AAY22805-Y22917 are derived from Human immunodeficiency virus  
 CC (HIV). Specifically, AAY22810-Y22910 are derived from gp41 proteins  
 CC of known HIV virus strains (AAY22810, AAY22871, AAY22888 and  
 CC AAY22903 represent consensus sequences of various sections of the gp41  
 CC protein). Sequences derived from the peptides are used to produce  
 CC constrained helical peptides of the invention. The constrained helical  
 CC peptide is produced by synthesizing an octapeptide in which both terminal  
 CC amino acids have a side-chain that includes a group able to form an amide  
 CC bond, and cyclizing the octapeptide by reacting the specified side-chain  
 CC residues with a difunctional linker to produce two amide bonds  
 CC The constrained helical peptides are used to treat or prevent HIV  
 CC infection, especially as vaccines that generate antibodies that  
 CC prevent viral membrane fusion or infectivity. Vaccines may contain  
 CC constrained helical peptides derived from several different strains of  
 CC HIV. The antibodies are also useful for diagnosing HIV infection. Other  
 CC uses for the constrained helical peptides are in affinity purification  
 CC of ligands (particularly where complete binding protein is not readily  
 CC available, e.g. replacements for protein A in immunoglobulin  
 CC purification); as epitope mimics for antibody production; for isolation  
 CC of synthetic antibody clones from phage display libraries, or as stable  
 CC forms of "floppy" peptides or proteins.  
 CC  
 SQ Sequence 269 AA;  
 QY 1 WIDIKQFVWY 10  
 DB 203 WFDITQWLWY 212  
 Query Match 48.4%; Score 45; DB 19; Length 269;  
 Best Local Similarity 60.0%; Pred. No. 69;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Search completed: March 26, 2003, 16:48:16  
 Job time : 30.1279 secs



GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 16:38:14 ; Search time 4.36047 Seconds  
(without alignments)  
142.679 Million cell updates/sec

Title: US-10-010-667A-21

Perfect score: 93  
Sequence: 1 WIDIKQFVWYPTPTF 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	93	100.0	339	1 STEA_HUMAN	O9uh88 homo sapien
2	52	55.9	458	1 AGLB_RHIME	O9z3i5 rhizobium m
3	45	48.4	865	1 ENV_HV1A.H	P04579 human immun
4	43.5	46.8	1167	1 WCL_NEUCR	O01371 neurospora
5	42	45.2	346	1 OPSF_ICTPU	O42266 ictalurus p
6	42	45.2	336	1 YOR1_YEAST	P46992 saccharomyc
7	42	45.2	685	1 ERF2_YEAST	P05453 saccharomyc
8	42	45.2	795	1 TLR1_MOUSE	O9epd1 mus musculu
9	42	45.2	847	1 ENV_HV1S1	P39560 human immun
10	41.5	44.6	224	1 YN85_YEAST	P51721 saccharomyc
11	41	44.1	405	1 WCAD_ECOLI	P12328 escherichia
12	41	44.1	430	1 PUCC_BACSU	O32140 bacillus su
13	41	44.1	703	1 ARYB_MANSB	P14297 manduca sex
14	41	44.1	843	1 ENV_HV1A2	P35961 human immun
15	41	44.1	854	1 ENV_SIVCZ	P17281 chimpanzee
16	41	44.1	856	1 ENV_HV1Z4	P05881 human immun
17	41	44.1	1509	1 MYSN_AACCA	P05659 acanthamoeb
18	41	44.1	1628	1 NAGH_CLOPE	P26831 clostridium
19	40	43.0	226	1 HAZP_RABIT	P20755 corycolagus
20	40	43.0	847	1 ENV_HV1A2	P05880 human immun
21	40	43.0	853	1 ENV_HV1A2	P12487 human immun
22	40	43.0	856	1 ENV_HV1A2	P04580 human immun
23	40	43.0	856	1 ENV_HV1A2	P05877 human immun
24	39.5	42.5	263	1 MPEI_LOLPR	P14946 lolium pere
25	39.5	42.5	263	1 MPEI_LOLPR	P14946 lolium pere
26	39.5	42.5	263	1 MPEI_LOLPR	P14946 lolium pere
27	39.5	42.5	263	1 MPEI_LOLPR	P14946 lolium pere
28	39.5	42.5	263	1 MPEI_LOLPR	P14946 lolium pere
29	39.5	42.5	263	1 MPEI_LOLPR	P14946 lolium pere
30	39.5	42.5	263	1 MPEI_LOLPR	P14946 lolium pere
31	39.5	42.5	263	1 MPEI_LOLPR	P14946 lolium pere
32	39.5	42.5	263	1 MPEI_LOLPR	P14946 lolium pere
33	39.5	42.5	263	1 MPEI_LOLPR	P14946 lolium pere

34	39	41.9	361	1 SERC_BACHD	O9kdm4 bacillus ha
35	39	41.9	459	1 NUDM_MACRO	P92668 macropus ro
36	39	41.9	847	1 MDH_ECOLI	P33137 escherichia
37	39	41.9	853	1 ENV_HV1A4	P04581 human immun
38	39	41.9	868	1 ENV_HV1A4	P05879 human immun
39	38.5	41.4	614	1 YEHQ_ECOLI	P33353 escherichia
40	38	40.9	258	1 RCER_RHOVI	P06008 rhodospirado
41	38	40.9	330	1 HEM2_AOUAR	O67876 aquilex aeo
42	38	40.9	566	1 MBHM_ECOLI	P37181 escherichia
43	38	40.9	702	1 DD4_MOUSE	O61496 mus musculu
44	38	40.9	707	1 J1P1_MOUSE	O9w4i9 mus musculu
45	38	40.9	708	1 J1P1_MOUSE	O9z237 r c-jun-aml

## ALIGNMENTS

RESULT 1	ID	STEA_HUMAN	STANDARD;	PRT;	339 AA.
AC	O9uh88	O95034;			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	15-JUN-2002	(Rel. 41, Last annotation update)			
DE	Six transmembrane epithelial antigen of prostate.				
GN	STEAP OR STEAP1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RN	SEQUENCE FROM N.A.				
RX	MEDLINE=20056277; PubMed=10588738;				
RA	Hubert R.S., Vivanco I., Chen E., Rastegar S., Leong K.,				
RA	Mitchell S.C., Madraswala R., Zhou Y., Kuo J., Raitano A.B.,				
RA	Jakobovits A., Salifian D.C., Afari D.E.H.,				
RT	"STEAP: A prostate-specific cell-surface antigen highly expressed in				
RT	human prostate tumors.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 96:14523-14528(1999).				
RL	[2]				
RP	SEQUENCE FROM N.A.				
RA	Abu-Threiden J., Stoneking T., Langston Y., Maupin R.,				
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Skin;				
RA	Strausberg R.;				
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.				
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).				
CC	-1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PROSTATE TUMORS.				
CC	-1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PROSTATE TUMORS.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; AF186249; AA017479.1; -				
DR	EMBL; AC005053; AAC79150.1; ALT_INIT.				
DR	EMBL; AC004969; AAD15620.1; ALT_INIT.				
DR	EMBL; BC011802; AAH1802.1; -				
DR	Genew; HGNC:11378; STEAP.				
DR	MIM; 604415; -				
KW	Transmembrane; Antigen.				
FT	TRANSMEM 71 91				POTENTIAL.
FT	TRANSMEM 119 139				POTENTIAL.
FT	TRANSMEM 164 184				POTENTIAL.
FT	TRANSMEM 218 238				POTENTIAL.
FT	TRANSMEM 258 278				POTENTIAL.
FT	TRANSMEM 291 311				POTENTIAL.
SQ	SEQUENCE 339 AA; 39851 MW; 55443A170C870387 CRC64;				



```

FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 367 367 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 403 403 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 407 407 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 625 625 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 634 634 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 825 825 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 865 AA; 97809 MW; 28828BC2314ADCAC CRC64;

Query Match 48.4%; Score 45; DB 1; Length 865;
Best Local Similarity 60.0%; Pred. No. 12;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 WIDKQFVWY 10
Db 681 WPDITQWLMY 690

RESULT 4
WCI_NEUCR STANDARD; PRT; 1167 AA.
ID WCI_NEUCR
AC Q01371;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE White collar 1 protein (WCI1).
WC-1.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxId=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RX MEDLINE=96203083; PubMed=8612589;
RA Ballario P., Vitorioso P., Magrelli A., Talora C., Cabibbo A.,
RA Macino G.;
RT "White collar-1, a central regulator of blue light responses in
RT Neurospora, is a zinc finger protein."
RL EMO J. 15:1650-1657(1996).
[2]
RP REVISIONS TO C-TERMINUS.
RA Ballario P.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR INVOLVED IN LIGHT
CC REGULATION. BINDS AND AFFECTS BLUE LIGHT REGULATION OF THE AL-3
CC GENE. WCI1 AND WC2 PROTEINS INTERACT VIA HOMOLOGOUS PAS DOMAINS,
CC BIND TO PROMOTERS OF LIGHT REGULATED GENES SUCH AS FRQ, AND
CC ACTIVATE TRANSCRIPTION.
CC -1- SUBUNIT: HETERODIMER OF WCI1 AND WC2 (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- INDUCTION: BY BLUE LIGHT.
CC -1- DOMAIN: THE GLUTAMINE-RICH DOMAIN MIGHT FUNCTION IN ACTIVATING
CC GENE EXPRESSION.
CC -1- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC -1- SIMILARITY: CONTAINS 3 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL; X94300; CAA63964.2; -.
DR HSSP; P17679; IGNF.
DR TRANSFAC; T02819; -.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS_domain.
DR InterPro; IPR00679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR Pfam; PF00785; PAC; 2.
DR Pfam; PF00989; PAS; 3.
DR SMART; SM00066; PAC; 2.
DR SMART; SM00091; PAS; 3.
DR SMART; SM00401; Znf_GATA; 1.
DR TIGRFAMs; TIGR00229; sensory_box; 3.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS00114; GATA_ZN_FINGER_2; 1.
DR PROSITE; PS00112; PAS; 3.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein; Repeat.
FT DOMAIN 16 61 GLN-RICH.
FT DOMAIN 381 452 PAS 1.
FT DOMAIN 469 508 PAC 1.
FT DOMAIN 574 644 PAS 2.
FT DOMAIN 650 691 PAC 2.
FT DOMAIN 693 763 PAS 3.
FT ZN_FING 934 959 GATA-TYPE.
FT DOMAIN 21 57 POLY-GLN.
FT DOMAIN 329 333 POLY-PRO.
SQ SEQUENCE 1167 AA; 127454 MW; 6489D04DAB50E838 CRC64;

Query Match 46.8%; Score 43.5; DB 1; Length 1167;
Best Local Similarity 66.7%; Pred. No. 27;
Matches 8; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY 3 DIKQFVWYTPPT 14
Db 529 DIQGYIW-TPPT 539

RESULT 5
OESP ICTPU STANDARD; PRT; 346 AA.
ID OESP ICTPU
AC Q42266;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 38, Last annotation update)
DE Parapinopsin.
OS Ictalurus punctatus (Channel catfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
OC Ictaluridae; Ictalurus.
OX NCBI_TaxId=7998;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=97477428; PubMed=9334384;
RA Blackshaw S., Snyder S.H.;
RT "Parapinopsin, a novel catfish opsin localized to the paraneal
RT organ, defines a new gene family."
RL J. Neurosci. 17:8083-8092(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: PARAPINEAL ORGAN.
CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC BE PHOSPHORYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL; AF028014; AAB84050.1; -.
DR HSBP; P02699; 1FDF.
DR InterPro; IPR000276; GPCR_Rhodopn.
DR InterPro; IPR001760; Opsin.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODOPSN.
DR PROSITE; PS00237; G_PROTEIN_RECP_F1_1; 1.
DR PROSITE; PS00262; G_PROTEIN_RECP_F1_2; 1.
DR PROSITE; PS00238; Opsin; 1.
KM Photoreceptor; Retinal protein; Transmembrane; Glycoprotein;
KW Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.
FT DOMAIN 1 29 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 30 54 1 (POTENTIAL).
FT TRANSMEM 55 66 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 67 91 2 (POTENTIAL).
FT TRANSMEM 92 106 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 107 126 3 (POTENTIAL).
FT TRANSMEM 127 145 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 146 169 4 (POTENTIAL).
FT TRANSMEM 170 193 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 194 221 5 (POTENTIAL).
FT DOMAIN 222 244 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 245 268 6 (POTENTIAL).
FT TRANSMEM 269 276 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 277 301 7 (POTENTIAL).
FT TRANSMEM 302 346 CYTOPLASMIC (POTENTIAL).
FT DISULFID 103 180 BY SIMILARITY.
FT BINDING 288 288 RETINAL CHROMOPHORE (BY SIMILARITY).
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 315 315 PALMITATE (BY SIMILARITY).
SQ SEQUENCE 346 AA; 38203 MW; A70871684F8FCTFD CRC64;

Query Match 45.2%; Score 42; DB 1; Length 346;
Best Local Similarity 66.7%; Pred. No. 14;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 7 FVWYTPPTF 15
| | | | |
Db 158 FVWTPPLP 166

RESULT 6
YXR1_YEAST STANDARD; PRT; 396 AA.
AC P46952;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 43.0 kDa protein in CP81-PP1 intergenic region.
GN YJL171C OR J0512.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Obermaier B., Piravandi E., Rinke M., Dondley H.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RL -1- SIMILARITY: TO YEAST YBR162C.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; Z49446; CA89466.1; -.
DR SGD; S0003707; YJL171C.
KW Hypothetical protein.

```

```

SQ SEQUENCE 396 AA; 43014 MW; 279B858E7512670A CRC64;

Query Match 45.2%; Score 42; DB 1; Length 396;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 IDIKQFVWYTPPTF 15
| | | | |
Db 83 LKIKQFAFYSPSGF 96

RESULT 7
ERF2_YEAST STANDARD; PRT; 685 AA.
ID P05453; P05420;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Eukaryotic peptide chain release factor 3 (ERF3) (ERF-3) (Omnipotent suppressor
DE protein 2) (G1 to S phase transition protein 1).
DE SUP35 OR SUP2 OR SUP12 OR GST1 OR SAL3 OR PNM2 OR YDR172W OR
GN YD9395.05.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=88329727; PubMed=3047009;
RA Kushnir V.V., Ter-Avanesyan M.D., Telchov M.V., Surguchov A.P.,
RA Smirnov V.N., Inge-Vechtomov S.G.;
RT "Nucleotide sequence of the SUP2 (SUP35) gene of Saccharomyces
RT cerevisiae.";
RL Gene 66:45-54(1988).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=87219095; PubMed=3556215;
RA Kushnir V.V., Ter-Avanesyan M.D., Surguchov A.P., Smirnov V.N.,
RA Inge-Vechtomov S.G.;
RT "Localization of possible functional domains in sup2 gene product of
RT the yeast Saccharomyces cerevisiae.";
RL FEBS Lett. 215:257-260(1987).
RN [3]
RP SEQUENCE FROM N.A.
RA MEDLINE=88172503; PubMed=3280807;
RA Wilson P.G., Culbertson M.R.;
RT "SUP12 suppressor protein of yeast. A fusion protein related to the
RT EF-1 family of elongation factors.";
RL J. Mol. Biol. 199:559-573(1988).
RN [4]
RP SEQUENCE FROM N.A.
RA MEDLINE=88296422; PubMed=2841115;
RA Kikuchi Y., Shimatake H., Kikuchi A.;
RT "A yeast gene required for the G1-to-S transition encodes a protein
RT containing an A-kinase target site and GTPase domain.";
RL EMBO J. 7:1175-1182(1988).
RN [5]
RP SEQUENCE FROM N.A.
RA STRAIN=S288C / AB972;
RA Murphy L., Harris D.B., Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RL [6]
RP FUNCTION.
RA MEDLINE=96016209; PubMed=7556078;
RA Stansfield I., Jones K.M., Kushnir V.V., Dagkesanandakaya A.R.,
RA Poznyakovskii A.I., Paushkin S.V., Nietras C.R., Cox B.S.,
RA Ter-Avanesyan M.D., Tuite M.F.;
RT "The products of the SUP5 (ERF1) and SUP35 genes interact to mediate
RT translation termination in Saccharomyces cerevisiae.";
RL EMBO J. 14:4365-4373(1995).
CC -1- FUNCTION: INVOLVED IN TRANSLATION TERMINATION. STIMULATES THE
CC ACTIVITY OF ERF1. BINDS GUANINE NUCLEOTIDES.

```

CC -1- SUBUNIT: HETERODIMER OF TWO SUBUNITS, ONE OF WHICH BINDS GTP.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.  
 CC ERK3 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M21129; AAA35133.1; -  
 DR EMBL; X07163; AAA30155.1; -  
 DR EMBL; Y00829; CAA68760.1; -  
 DR EMBL; Z46727; CAA6677.1; -  
 DR PIR; S00733; EFBS2.  
 DR SGD; S0002579; SUP35.  
 DR InterPro; IPR004160; EFTU Cterm.  
 DR InterPro; IPR004161; EFTU D2.  
 DR InterPro; IPR00795; EF GTPbind.  
 DR Pfam; PF00009; GTP\_EFTU\_1.  
 DR Pfam; PF03143; GTP\_EFTU\_D3; 1.  
 DR Pfam; PF03144; GTP\_EFTU\_D2; 1.  
 DR PROSITE; PS00301; EFATOR GTP; 1.  
 DR Protein biosynthesis; Repeat; Phosphorylation; GTP-binding.  
 FT DOMAIN 5 135  
 FT SIMILAR 139 249  
 FT NP\_BIND 254 685  
 FT NP\_BIND 267 274  
 FT NP\_BIND 344 348  
 FT NP\_BIND 406 409  
 FT ACT\_SITE 273 273  
 FT ACT\_SITE 407 407  
 FT MOD\_RES 341 341  
 FT CONFLICT 53 53  
 FT SEQUENCE 685 AA; 76551 MW; 43912ACD77DFA153 CRC64;  
 SQ  
 Query Match 45.2%; Score 42; DB 1; Length 685;  
 Beel Local Similarity 53.8%; Pred. No. 28;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 Oy 2 IDIKOFWYTPPT 14  
 Db 459 VDPKCPWTGPT 471  
 RESULT 8  
 ID TLR1\_MOUSE STANDARD; PRT; 795 AA.  
 AC Q9EPQ1; Q9EPW5; 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Toll-like receptor 1 precursor (Toll/interleukin-1 receptor-like)  
 GN (TIR).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Macrophage;  
 RX MEDLINE=20558581; PubMed=11095740;  
 RA Ozinsky A., Underhill D.M., Fontenot J.D., Hajjar A.M., Smith K.D.,  
 RA Wilson C.B., Schroeder L., Aderem A.;  
 RT "The repertoire for pattern recognition of pathogens by the innate  
 RT immune system is defined by cooperation between Toll-like  
 RT receptors";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:13766-13771(2000).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Macrophage;  
 RX MEDLINE=20571875; PubMed=11123271;  
 RA Hajjar A.M., O'Mahony D.S., Ozinsky A., Underhill D.M., Aderem A.,  
 RA Klebanoff S.J., Wilson C.B.;  
 RT "Functional interactions between Toll-like receptor (TLR) 2 and TLR1  
 RT or TLR6 in response to phenol-soluble modulin";  
 RL J. Immunol. 166:15-19(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Spleen;  
 RA Thomson D.P., Campbell C.C., Liew F.Y., Xu D.;  
 RT "Cloning of Mus musculus Toll-like receptor 1";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Participates in the innate immune response to microbial  
 CC agents. Cooperates with TLR2 and modulates the response to  
 CC microbial constituents. Acts via MyD88 and TRAF6, leading to NF-  
 CC kappa-B activation, cytokine secretion and the inflammatory  
 CC response (By similarity).  
 CC -1- SUBUNIT: Binds TLR2 via their respective extracellular domains.  
 CC Binds MyD88 via their respective TIR domains (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Plasma membrane and  
 CC phagosomes.  
 CC -1- SIMILARITY: BELONGS TO THE TOLL-LIKE RECEPTOR FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 TIR DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 9 LEUCINE-RICH REPEATS (LRR).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AY009154; AAG37302.1; -  
 DR EMBL; AF316985; AAG35062.1; -  
 DR HSSP; O60603; IFYW.  
 DR MGD; MGI:1341295; Tlr1.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR000483; LRR Cterm.  
 DR InterPro; IPR003592; LRR out.  
 DR InterPro; IPR003591; LRR typ.  
 DR InterPro; IPR000157; TIR domain.  
 DR Pfam; PF00560; LRR; 8.  
 DR Pfam; PF01463; LRRCT; 1.  
 DR Pfam; PF01582; TIR; 1.  
 DR PRINTS; PR00019; LEURICHRPT.  
 DR SMART; SM00370; LRR; 3.  
 DR SMART; SM00082; LRRCT; 1.  
 DR SMART; SM00369; LRR\_TYP; 6.  
 DR SMART; SM00255; TIR; 1.  
 DR PROSITE; PS50104; TIR; 1.  
 DR Receptor; Immune response; Inflammatory response; Signal;  
 KW Transmembrane; Repeat; Leucine-rich repeat; Glycoprotein.  
 FT SIGNAL 1 25  
 FT CHAIN 26 795  
 FT DOMAIN 26 582  
 FT TRANSMEM 583 603  
 FT DOMAIN 604 795  
 FT REPEAT 46 70  
 FT REPEAT 71 94  
 FT REPEAT 96 116  
 FT REPEAT 117 140  
 FT REPEAT 140 140  
 FT REPEAT 374 397  
 FT REPEAT 400 427  
 FT REPEAT 447 470  
 FT REPEAT 471 494  
 FT REPEAT 496 515  
 FT DOMAIN 638 782  
 FT CARBOHYD 38 38  
 FT CARBOHYD 59 59  
 FT CARBOHYD 88 88  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 266 266 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 88 N -> D (IN REF. 3).
SQ SEQUENCE 795 AA; 90672 MW; 855356429872D232 CRC64;

Query Match
Best Local Similarity 45.2%; Score 42; DB 1; Length 795;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 1 WIDIKOPVWYTP 12
Db 264 FINILOIWHPT 275

RESULT 9
ENV_HV1S1 STANDARD; PRT; 847 AA.
AC P19550;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Envelope polypeptide GP160 precursor (Contains: Exterior membrane glycoprotein (GP120); Transmembrane glycoprotein (GP41)).
GN ENV.
OS Human immunodeficiency virus type 1 (SF162 isolate) (HIV-1).
OC Viruses; Retroviridae; Retroviridae; Lentivirus.
OX NCBI_Taxid=11691;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90347835; PubMed=2384920;
RA Cheng-Mayer C., Quiraga M., Tung J.W., Levy J.;
RT "Viral determinants of human immunodeficiency virus type 1 T-cell or macrophage tropism, cytopathogenicity, and CD4 antigen modulation.";
RL J. Virol. 64:4390-4396(1990).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M65024; AAA45072.1; -.
DR HIV; M38428; ENVSR162.
DR InterPro; IPR000328; Env GP41.
DR InterPro; IPR00077; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW Aids; Coat protein; Polypeptide; Glycoprotein; Transmembrane;
KW signal.
FT SIGNAL 1 29
FT CHAIN 30 502 EXTERIOR MEMBRANE GLYCOPROTEIN.
FT CHAIN 503 847 TRANSMEMBRANE GLYCOPROTEIN.
FT DISULFID 53 73 BY SIMILARITY.
FT DISULFID 118 203 BY SIMILARITY.
FT DISULFID 125 194 BY SIMILARITY.
FT DISULFID 130 155 BY SIMILARITY.
FT DISULFID 216 245 BY SIMILARITY.
FT DISULFID 226 237 BY SIMILARITY.
FT DISULFID 294 328 BY SIMILARITY.
FT DISULFID 374 435 BY SIMILARITY.
FT DISULFID 381 408 BY SIMILARITY.
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 154 154 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 195 195 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 274 274 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 438 438 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 454 454 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 602 602 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 607 607 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 847 AA; 96135 MW; 0A901317FD7FF2AB CRC64;

Query Match
Best Local Similarity 45.0%; Score 42; DB 1; Length 847;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 1 WIDIKOPVWY 10
Db 663 WFDISKMLWY 672

RESULT 10
YN89_YEAST STANDARD; PRT; 224 AA.
ID YN89_YEAST
AC P53721;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 25.3 kDa protein in YJM2-ARE2 intergenic region.
GN YNR018W OR N3185.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_Taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Pohl T.M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z71633; CAA96297.1; -.
DR SGD; S0005301; YNR018W.
DR Hypothetical protein; Transmembrane.
KW TRANSMEM 19 39 POTENTIAL.
FT TRANSMEM 50 70 POTENTIAL.
FT TRANSMEM 152 172 POTENTIAL.
FT DOMAIN 196 199 POLY-GLU.
SQ SEQUENCE 224 AA; 25344 MW; FA2C528A008CFE7C CRC64;

Query Match
Best Local Similarity 44.6%; Score 41.5; DB 1; Length 224;
Matches 8; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

Oy 1 WIDIKOPVWYTPPT 14
Db 52 W-SIKTALWITPPT 64

```

```

RESULT 11
ID WCAD_ECOLI STANDARD; PRT; 405 AA.
AC P71238; P76385;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative colanic acid polymerase.
GN WCAD OR B2056.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxId=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=96326333; PubMed=8759852;
RA Stevenson G., Andrianopoulos K., Hobbs M., Reeves P.R.;
RT "Organization of the Escherichia coli K-12 gene cluster responsible
for production of the extracellular polysaccharide colanic acid.";
RL J. Bacteriol. 178:4885-4893(1996).
RN [2]
RP REVISIONS.
RC STRAIN=K12;
RA Reeves P.R.;
RT Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP -1- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF THE SLIME POLYSACCHARIDE
COLANIC ACID.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U38473; AAC77839.1; -
DR EMBL; AE000295; AAC75117.1; -
DR Ecogene; BG13572; wcad.
KW Lipopolysaccharide biosynthesis; Complete proteome.
FT CONFLICT 108 P -> A (IN REF. 2).
SQ SEQUENCE 405 AA; 45409 MW; A3D9D91255686043 CRC64;

Query Match 44.1%; Score 41; DB 1; Length 405;
Best Local Similarity 57.1%; Pred. No. 24;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```

```

GN PUCK.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxId=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bartero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boutsier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Enliant K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi N.,
RA Giuseppe G., Guy B.J., Haga K., Haeche J., Harwood C.R., Henaut A.,
RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Ilaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koester P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Moone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Prescarn E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier P., Vassaretto A.,
RA Viari A., Wambit R., Wedler E., Wedler H., Weitzenger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein B., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis.";
RL Nature 390:249-256(1997).
RN [2]
RP FUNCTION.
RX PubMed=11344136;
RA Schultze A.C., Nygaard P., Saxild H.H.;
RT "Functional analysis of 14 genes that constitute the purine catabolic
RT pathway in Bacillus subtilis and evidence for a novel region
RT controlled by the Puck transcription activator.";
RL J. Bacteriol. 183:3293-3302(2001).
CC -----
CC -1- FUNCTION: Uptake of uric acid.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE XANTHINE/URACIL PERMEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z99120; CAB15234.1; -
DR Subtilist; BG13985; Puck.
DR InterPro: IPR000444; Xan ur permease.
DR Pfam: PF00860; xan ur permease; 1.
DR TIGRPFAM; TIGR00801; ncr2; 1.
DR PROSITE; PS01116; XANTH_URACIL_PERMASE; 1.
KW Transmembrane; Transport; Complete proteome.
FT TRANSMEM 18
FT TRANSMEM 38
FT TRANSMEM 43
FT TRANSMEM 63
FT TRANSMEM 67
FT TRANSMEM 87
FT TRANSMEM 97
FT TRANSMEM 117
FT TRANSMEM 122
FT TRANSMEM 142
FT TRANSMEM 163
FT TRANSMEM 185
FT TRANSMEM 205
FT TRANSMEM POTENTIAL.
FT TRANSMEM POTENTIAL.
FT TRANSMEM POTENTIAL.

```

```

FT TRANSMEM 209 229 POTENTIAL.
FT TRANSMEM 233 253 POTENTIAL.
FT TRANSMEM 274 294 POTENTIAL.
FT TRANSMEM 310 330 POTENTIAL.
FT TRANSMEM 333 353 POTENTIAL.
FT TRANSMEM 369 389 POTENTIAL.
SQ SEQUENCE 430 AA; 44861 MW; 9F3D18D10A0572F CRC64;

Query Match
Best Local Similarity 33.3%; Score 41; DB 1; Length 430;
Matches 5; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 WIDKQFWMYTPPTF 15
Db 215 WLVHPSLFYCGPPTF 229

RESULT 13
ARYB MANSE
ID ARYB MANSE STANDARD; PRT; 703 AA.
AC P14257;
DT 01-JUN-1990 (Rel. 13, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Arylphorin beta subunit precursor.
OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC Insecta; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Diptera; Sphingidae; Sphingidae; Sphinginae; Manduca.
OX NCBI_Taxid=7130;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90037032; PubMed=2808410;
RA Willott E., Wang X.-Y., Wells M.A.;
RT "cDNA and gene sequence of Manduca sexta arylphorin, an aromatic
RT amino acid-rich larval serum protein. Homology to arthropod
RT hemocyanins."
RL J. Biol. Chem. 264:19052-19059(1989).
CC -1- FUNCTION: ARYLPHORIN IS A LARVAL STORAGE PROTEIN (LSP) WHICH MAY
CC SERVE AS A STORAGE PROTEIN USED PRIMARILY AS A SOURCE OF AROMATIC
CC AMINO ACIDS FOR PROTEIN SYNTHESIS DURING METAMORPHOSIS. IT IS A
CC CONSTITUENT OF THE SCLEROTIZING SYSTEM OF THE CUTICLE, AND SERVES
CC AS A CARRIER FOR ECDYSTEROID HORMONE.
CC -1- SUBUNIT: ARYLPHORIN IS A HEXAMER OF SUBUNITS ALPHA AND BETA.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: FAT BODY.
CC -1- SIMILARITY: TO ARYB, TO B.MORI STORAGE PROTEINS 1 AND 2, AND TO
CC ARTHROPOD HEMOCYANINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M28395; AAA29304.1; -;
DR EMBL; M28397; AAA29305.1; -;
DR PIR; B34434; B34434.
DR HSSP; P04253; 10XY.
DR InterPro; IPR000866; Hemocyanin.
DR InterPro; IPR005203; hemocyanin_C.
DR InterPro; IPR005204; hemocyanin_N.
DR Pfam; PF00372; hemocyanin_1.
DR Pfam; PF03722; hemocyanin_N; 1.
DR Pfam; PF03723; hemocyanin_C; 1.
DR PRINTS; PRO0187; HEMOCYANIN.
DR PROSITE; PS00209; HEMOCYANIN_1; 1.
DR PROSITE; PS00210; HEMOCYANIN_2; 1.
KW Signal; Storage protein; Glycoprotein; Multigene family.

```

```

FT SIGNAL 1 16
FT CHAIN 17 703 ARYLPHORIN BETA SUBUNIT.
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 703 AA; 83848 MW; 24B7DBB60DD2E3FA CRC64;

Query Match
Best Local Similarity 60.0%; Score 41; DB 1; Length 703;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 DIKQFWMYTP 12
Db 287 DIPEFSWYSP 296

RESULT 14
ENV_HVIY2
ID ENV_HVIY2 STANDARD; PRT; 843 AA.
AC P35961;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Envelope glycoprotein GP160 precursor [Contains: Exterior membrane
DE glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
GN ENV.
OS Human immunodeficiency virus type 1 (YU-2 isolate) (HIV-1).
OC Viruses; Retroviridae; Retroviridae; Lentivirus.
OX NCBI_Taxid=36377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93021387; PubMed=1404605;
RA Li Y., Hui H., Burgess C.J., Price R.W., Sharp P.M., Hahn B.H.,
RA Shaw G.M.;
RT "Complete nucleotide sequence, genome organization, and biological
RT properties of human immunodeficiency virus type 1 in vivo: evidence
RT for limited defectiveness and complementation."
RL J. Virol. 66:6587-6600(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M93258; -; NOT_ANNOTATED_CDS.
DR PIR; H44001; H44001.
DR InterPro; IPR000328; Env GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW Aids; Coat protein; Polypeptide; Glycoprotein; Transmembrane;
KW Signal.
FT SIGNAL 1 29
FT CHAIN 30 489 EXTERIOR MEMBRANE GLYCOPROTEIN.
FT CHAIN 490 843 TRANSMEMBRANE GLYCOPROTEIN.
FT TRANSMEM 738 755 POTENTIAL.
FT DISULFID 53 73 BY SIMILARITY.
FT DISULFID 118 201 BY SIMILARITY.
FT DISULFID 125 192 BY SIMILARITY.
FT DISULFID 130 155 BY SIMILARITY.
FT DISULFID 214 243 BY SIMILARITY.
FT DISULFID 224 235 BY SIMILARITY.
FT DISULFID 292 326 BY SIMILARITY.
FT DISULFID 373 405 BY SIMILARITY.
FT DISULFID 380 405 BY SIMILARITY.
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 154 154 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 158 158 N-LINKED (GLCNAC. . .) (POTENTIAL).

```



FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 193 193 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 230 230 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 237 237 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 327 327 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 389 389 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 395 395 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 435 435 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 450 450 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 612 612 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 624 624 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 803 803 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 843 AA, 95648 MW, C69DFD971C918B71 CRC64;

Query Match 44.1%; Score 41; DB 1; Length 843;  
 Best Local Similarity 50.0%; Pred. No. 50;  
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 WIDIKQFWY 10  
 Db 659 WPDITKMLWY 668

FT CHAIN 501 854 TRANSMEMBRANE GLYCOPROTEIN.  
 FT TRANSMEM 501 517 POTENTIAL.  
 FT TRANSMEM 675 693 POTENTIAL.  
 FT TRANSMEM 805 821 POTENTIAL.  
 FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 143 143 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 154 154 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 158 158 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 195 195 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 229 239 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 250 260 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 257 267 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 274 274 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 384 384 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 426 426 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 446 446 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 450 450 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 601 601 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 608 608 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 854 AA, 95803 MW, 2E249AFAD4F2D9B3 CRC64;

Query Match 44.1%; Score 41; DB 1; Length 854;  
 Best Local Similarity 50.0%; Pred. No. 51;  
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 WIDIKQFWY 10  
 Db 663 WPDITKMLWY 672

Search completed: March 26, 2003, 16:45:09  
 Job time : 6.36047 secs

RESULT 15  
 ENV\_SIVCZ STANDARD; PRT; 854 AA.  
 ID ENV\_SIVCZ P17281;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Envelope polypeptide GP160 precursor [Contains: Exterior membrane glycoprotein (GP120); Transmembrane glycoprotein (GP41)].  
 GN ENV.  
 OS Chimpanzee immunodeficiency virus (SIV(CPZ)) (CIV).  
 OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11723;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90259077; PubMed=2180136;  
 RA Huet T., Cheylier R., Meyerhans A., Roelants G., Main-Hobson S.;  
 RT "Genetic organization of a chimpanzee lentivirus related to HIV-1."  
 RL Nature 345:356-359(1990).  
 CC -1- SIMILARITY: STRONG, WITH HIV-1 ENV POLYPROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X52154; CAA36407.1; -.  
 DR FIR; S08990; VCLJST.  
 DR HIV; X52154; ENVSCPZ.  
 DR InterPro; IPR000328; ENV GP41.  
 DR Pfam; PF00516; GP120; 1.  
 DR Pfam; PF00517; GP41; 1.  
 KW AIDS; Coat protein; Polypeptide; Glycoprotein; Transmembrane; Signal.  
 FT SIGNAL 1 30  
 FT CHAIN 31 500 EXTERIOR MEMBRANE GLYCOPROTEIN.



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 16:38:14 ; Search time 9.88372 Seconds

(without alignments)  
142.679 Million cell updates/sec

Title: US-10-010-667a-20

Perfect score: 192  
Sequence: 1 RRSYRYKLIMWAVYQVQNKEDAWIEHDVWMEI 34

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	192	100.0	339	1 STEA_HUMAN	Q9ubh8 homo sapien
2	58.5	30.5	1418	1 CELL_CAEEL	P34641 caenorhabdi
3	58	30.2	900	1 GGAB_BACSU	P46918 bacillus su
4	54	28.1	630	1 Y4BJ_RHISN	P55377 rhizobium s
5	52	27.1	79	1 YOKC_BACSU	P54566 bacillus s
6	52	27.1	1813	1 UN13_CAEEL	P27715 caenorhabdi
7	51	26.6	495	1 GLPK_SYNY3	P74260 synechocyst
8	51	26.6	1017	1 ACAD_ARATH	Q911k7 arabidopsis
9	50	26.0	784	1 ALP4_SCHPO	Q91705 schizosacch
10	50	26.0	1574	1 SYJI_RAT	O62910 rattus norv
11	50	26.0	5430	1 ACP7_HUMAN	O9upn3 homo sapien
12	49.5	25.8	95	1 YQ04_BACAN	Q9rn78 bacillus an
13	49.5	25.8	274	1 DAPD_BUCAI	P57333 buchnera ap
14	49.5	25.8	350	1 YC73_METJA	O58669 methanococc
15	49.5	25.8	940	1 SYV_CHLPS	Q52967 chlamydia p
16	49	25.5	512	1 Y4W4_RHISN	P55679 rhizobium s
17	49	25.5	837	1 NCM2_HUMAN	O15394 homo sapien
18	49	25.5	837	1 NCM2_MOUSE	O35136 mus musculu
19	49	25.5	1324	1 SYJI_BOVIN	O18964 bos taurus
20	49	25.5	1575	1 SYJI_HUMAN	O43426 homo sapien
21	48.5	25.3	266	1 IF2A_SULSO	Q97279 sulfolobus
22	48.5	25.3	339	1 MOZL_CAEEL	Q9tzm2 caenorhabdi
23	48	25.0	266	1 YAV8_SCHPO	O10216 schizosacch
24	48	25.0	864	1 AGLU_MUCA	Q92442 mucor javan
25	48	25.0	942	1 ENV_CAEVG	P11627 caprine arc
26	47.5	24.7	561	1 O16G_BACSU	O06994 bacillus su
27	47.5	24.7	796	1 DHG_ECOLI	P15877 escherichia
28	47.5	24.7	808	1 DHG_GLUOX	P27175 gluconobact
29	47.5	24.7	1103	1 CYGF_BOVIN	O02740 bos taurus
30	47	24.5	244	1 CYBH_BRAJA	P21960 bradyrhizob
31	47	24.5	586	1 VATA_HALVO	Q48332 halobacteri
32	47	24.5	675	1 VP55_YEAST	Q92331 saccharomyc
33	47	24.5	1002	1 POL_HV1U4	P24740 human immun

34	47	24.5	1133	1 ATX9_TETTR	O95050 tetrahymena
35	47	24.5	1503	1 TRU2_HUMAN	O94759 homo sapien
36	46.5	24.2	215	1 CIB2_HUMAN	O60939 homo sapien
37	46.5	24.2	215	1 CIB2_RAT	P54900 rattus norv
38	46.5	24.2	463	1 SYN_GLOAB	O97656 clostridium
39	46.5	24.2	508	1 O16G_BACSP	P29093 bacillus sp
40	46.5	24.2	555	1 O16G_BACCO	O45101 bacillus co
41	46	24.0	220	1 DUB7_MOUSE	O9qy18 mus musculu
42	46	24.0	236	1 BIR8_GORGO	O95m71 gorilla gor
43	46	24.0	236	1 BIR8_HUMAN	O95p09 homo sapien
44	46	24.0	236	1 BIR8_PANTR	O95n72 pan troglod
45	46	24.0	242	1 DUB6_MOUSE	O54946 mus musculu

## ALIGNMENTS

RESULT 1	STEA_HUMAN	STANDARD	PRT	339 AA.
ID	STEA_HUMAN			
AC	Q9ubh8; Q95034;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Six transmembrane epithelial antigen of prostate.			
GN	STEAP OR STEAP1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20056277; PubMed=1058738;			
RA	Hubert R.S., Vivanco I., Chen E., Rastegar S., Leong K.,			
RA	Mitchell S.C., Madraswala R., Zhou Y., Kuo J., Raitano A.B.,			
RA	Jakovlevs A., Saffran D.C., Afar D.E.H.,			
RT	"STEAP: a prostate-specific cell-surface antigen highly expressed in			
RT	human prostate tumors."			
RL	Proc. Natl. Acad. Sci. U.S.A. 96:14523-14528(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Ku-Threiden U., Stoneking T., Langston Y., Maupin R.,			
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Skin;			
RA	Strausberg R.,			
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
CC	- SUBCELLULAR LOCATION: Integral membrane protein (Potential).			
CC	- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PROSTATE TUMORS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and that this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
CC	-----			
DR	EMBL; AF186249; AAF17479.1; -			
DR	EMBL; AC005053; AAC79150.1; ALT_INIT.			
DR	EMBL; AC004969; AAD15620.1; ALT_INIT.			
DR	EMBL; BC011802; AAD11802.1; -			
DR	Genew; HGNC:11378; STEAP.			
DR	MTM; 60415; -			
KW	Transmembrane; Antigen.			
FT	TRANSMEM 71			POTENTIAL.
FT	TRANSMEM 91			POTENTIAL.
FT	TRANSMEM 119			POTENTIAL.
FT	TRANSMEM 164			POTENTIAL.
FT	TRANSMEM 218			POTENTIAL.
FT	TRANSMEM 258			POTENTIAL.
FT	TRANSMEM 291			POTENTIAL.
FT	TRANSMEM 311			POTENTIAL.
SQ	SEQUENCE 339 AA; 39851 MW; 55443A170C870387 CRC64;			

Query Match 100.0%; Score 192; DB 1; Length 339;  
 Best Local Similarity 100.0%; Pred. No. 9, 4e-19;  
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RRSRYKLNMAYQVQONKEDAMIEHDVNRMEI 34  
 Db 185 RRSRYKLNMAYQVQONKEDAMIEHDVNRMEI 218

## RESULT 2

CELL CAEEL STANDARD; PRT; 1418 AA.

AC P34641;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ccd-11 protein.  
 GN Ccd-11 OR ZK512.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 NC NCB1\_TaxId=6239;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Alincough R., Anderson K., Baynes C., Berks M.,  
 Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,  
 Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,  
 Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 Johnston L., Jones M., Kershaw J., Kitzens J., Lalister N.,  
 Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,  
 Sims M., Smailson N., Smith A., Smith M., Sonhammer E., Staden R.,  
 Suleston J., Thierry-Mieg J., Thomas K., Vaadin M., Vaughan K.,  
 Waterston R., Watson A., Weinstock L., Wilkinson-Spratt J.,  
 Woildman P.;  
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans";  
 RL Nature 368:32-38(1994).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: Z22177; CAAB0145.1; -  
 DR PIR: S40764; S40764.  
 DR WormPep: ZK512.3; CE00409.  
 DR InterPro: IPR002111; Cat\_channel\_TpL.  
 SQ SEQUENCE 1418 AA; 159237 MW; 4FC83B9E7ADF7788 CRC64;

Query Match 30.5%; Score 58.5; DB 1; Length 1418;  
 Best Local Similarity 43.5%; Pred. No. 5.2;  
 Matches 10; Conservative 5; Mismatches 5; Indels 3; Gaps 1;

Qy 15 QVOONKED--AWIEHDVNRMEI 34  
 Db 1355 EVANENDTDNAWTEHDVNRMEI 1377

RESULT 3  
 GGAB\_BACSU STANDARD; PRT; 900 AA.

AC P46978;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Minor teichoic acids biosynthesis protein ggab.

GN GGAB.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 NC NCB1\_TaxId=1423;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunz F., Ogatawara N., Moszer I., Albertini A.M., Alloni G.,  
 Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 Brouillet S., Bruschi C.V., Caldwell B., Capano V., Carter N.M.,  
 Choi S.K., Codani U.J., Conneron I.F., Cummings N.J., Daniel R.A.,  
 Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 Entlian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,  
 Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
 Hilbert H., Holstappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 Joris B., Karamata D., Kasahara Y., Klier-Blanchard M., Klein C.,  
 Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,  
 Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,  
 Paro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M.,  
 Pressac A.E., Puig P., Portelle D., Rapoport G., Rey M., Reynolds S.,  
 Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y.,  
 Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 Sekiguchi J., Sekowska A., Seir S.J., Serron P., Shin B.S., Soldo B.,  
 Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 Tosato V., Uchiyama S., Vandendol M., Vanlier F., Vasseroiti A.,  
 Viari A., Wambutt R., Wedler E., Wedler H., Wetzenecker T.,  
 Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,  
 Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa K., Danchin A.,  
 RA "The complete genome sequence of the Gram-positive bacterium Bacillus  
 subtilis";  
 RL Nature 390:249-256(1997).

CC -1- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF GALACTOSAMINE-CONTAINING  
 CC MINOR TEICHOIC ACIDS.  
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U13979; AAA7513.1; -  
 DR EMBL: Z59122; CAB15565.1; -  
 DR Subtilisin, BG11192; ggab.  
 DR InterPro: IPR001173; Glycosyltransferase\_2.  
 DR Pfam: PF00535; Glycosyltransferase; Complete proteome.  
 DR Transference; Glycosyltransferase; Complete proteome.  
 SQ SEQUENCE 900 AA; 107154 MW; FA66495488C262F CRC64;

Query Match 30.2%; Score 58; DB 1; Length 900;  
 Best Local Similarity 33.3%; Pred. No. 3.7;  
 Matches 11; Conservative 8; Mismatches 14; Indels 0; Gaps 0;

Qy 2 RSYRYKLNMAYQVQONKEDAMIEHDVNRMEI 34  
 Db 591 RSRHKLTLTSSKRVISTHADIWVNRPFMEI 623

```

RESULT 4
Y4BU RHISN STANDARD; PRT; 630 AA.
AC P55377;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 67.9 kDa protein Y4BU.
GN Y4BU.
OS Rhizobium sp. (strain NGR234).
OC Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -1- SIMILARITY: DISTANTLY RELATED TO PEPTIDASE FAMILY 82C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000066; AAB91625.1; -
DR MEROPS; S01.UPC; -
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
KW Hypothetical protein; Plasmid; Transmembrane.
FT TRANSMEM 8
FT TRANSMEM 258 278
FT TRANSMEM 258 278
FT TRANSMEM 258 278
SQ SEQUENCE 630 AA; 67854 MW; 4C1FB7BA2B844A02 CRC64;

Query Match 28.1%; Score 54; DB 1; Length 630;
Best Local Similarity 35.5%; Pred. No. 8.5;
Matches 11; Conservative 5; Mismatches 7; Indels 8; Gaps 1;

QY 1 RRSYRYKLMMAYQVOQNKEDAWIEHDVWR 31
DB 296 RRYRRLFGWEYR-----AKLEPEIWR 318

RESULT 5
YOKC_BACSU STANDARD; PRT; 79 AA.
AC P54566;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein YOKC.
GN YOKC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=168 / JH642;
RX MEDLINE=97124195; PubMed=8969508;
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,
RA Kobayashi Y.;
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
RT the Bacillus subtilis genome containing the skin element and many
RT sporulation genes.";
RL Microbiology 142:3103-3111(1996).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Bouteier L., Brans A., Braum M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conterton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devigne K.M., Dusterhoff A., Ehrlich S.D., Emerson P.T.,
RA Ertan K.D., Errington J., Fabre C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Gim S.Y., Glaser P., Goffeau A., Golightly B.O., Grandi G.,
RA Guisepi G., Guy B.D., Haga K., Haiech J., Harwood C.R., Henauf A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blandhard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Prescan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpersa P., Tognoni A.,
RA Toaslo V., Uchiyama S., Vandenbol M., Vannier P., Vassarotti A.,
RA Viati A., Wambut R., Wedler E., Wedler H., Weltzenegger T.,
RA Winters P., Wipet A., Yamamoto H., Yamane K., Yasunaga K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zumbstein A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D84432; BAA12635.1; -
DR EMBL; Z99116; CAB14297.1; -
DR Subtilist; BG11758; YOKC.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 79 AA; 9287 MW; 34852BDAEA4805D2 CRC64;

Query Match 27.1%; Score 52; DB 1; Length 79;
Best Local Similarity 28.1%; Pred. No. 1.6;
Matches 9; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

QY 2 RSYRYKLMMAYQVOQNKEDAWIEHDVWR 33
DB 23 QNHPTLLHMSIGAESIKKDWLQDENTFE 54

RESULT 6
UN13_CABEL STANDARD; PRT; 1813 AA.
AC P27715; OL7665; Q23512;
DT 01-AUG-1992 (Rel. 23, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Photol ester/diacylglycerol-binding protein unc-13.
GN UNC-13 OR ZK524.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).

```

RX STRAIN-Bristol N2;  
 RX MEDLINE=91298538; PubMed=2062851;  
 RA Maruyama I.N., Brenner S.;  
 RT "A phorbol ester/diacylglycerol-binding protein encoded by the unc-13  
 gene of *Caenorhabditis elegans*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:5729-5733(1991).  
 RN [2]  
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1 AND 3).  
 RC STRAIN-Bristol N2;  
 RA Gardner A.E., Lloyd C.;  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP ZINC AND PHORBOL-ESTERS BINDING.  
 RX MEDLINE=91075060; PubMed=1445255;  
 RA Ahmed S., Maruyama I.N., Kozma R., Lee J., Brenner S., Lim L.;  
 RT "The *Caenorhabditis elegans* unc-13 gene product is a phospholipid-  
 dependent high-affinity phorbol ester receptor.";  
 RL Biochem. J. 287:995-999(1992).  
 CC -1- FUNCTION: MAY FORM PART OF A SIGNAL TRANSDUCTION PATHWAY,  
 CC TRANSDUCING THE SIGNAL FROM DIACYLGLYCEROL TO EFFECTOR  
 CC FUNCTIONS. ONE SUCH FUNCTION COULD BE THE RELEASE OF  
 CC NEUROTRANSMITTER FROM NEURONS.  
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms, 1/ZK524.2a (shown here), 2, 3  
 CC and 4/ZK524.2b; may be produced by alternative splicing.  
 CC -1- DOMAIN: THE PHORBOL ESTER BINDING ACTIVITY IS ZINC AND CALCIUM-  
 CC DEPENDENT.  
 CC -1- MISCELLANEOUS: MUTATIONS IN UNC-13 CAUSE DIVERSE NERVOUS SYSTEM  
 CC DEFECTS.  
 CC -1- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG  
 CC BINDING DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 3 C2 DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M62830; AAA93094.1; -;  
 DR EMBL; U50735; AAA9780.1; -;  
 DR EMBL; U50735; AAA9781.1; -;  
 DR EMBL; Z92779; CAB07173.1; -;  
 DR EMBL; Z79694; CAB07173.1; JOINED.  
 DR EMBL; Z73912; CAB07173.1; JOINED.  
 DR EMBL; Z73912; CAB98146.1; -;  
 DR EMBL; Z73912; CAB98147.1; -;  
 DR EMBL; Z79694; CAB98147.1; JOINED.  
 DR EMBL; Z92779; CAB98147.1; JOINED.  
 DR EMBL; Z79694; CAB01866.1; JOINED.  
 DR EMBL; Z92779; CAB01866.1; JOINED.  
 DR EMBL; Z73912; CAB01866.1; JOINED.  
 DR PIR; A41101; A41101.  
 DR HSSP; P28867; IPTO.  
 DR WormRep; ZK524.2a; CE15371.  
 DR WormRep; ZK524.2b; CE15287.  
 DR InterPro; IPR000008; C2.  
 DR InterPro; IPR002219; DAG\_pe-bind.  
 DR Pfam; PF00130; DAG\_pe-bind; 1.  
 DR Pfam; PF00168; C2; 3.  
 DR PRINTS; PR00360; C2DOMAIN.  
 DR PRINTS; PR00008; DAGPEDOMAIN.  
 DR SMART; SMO0109; C1; 1.

DR SMART; SMO0239; C2; 3.  
DR PROSITE; PSS00499; C2 DOMAIN 1; 1.  
DR PROSITE; PSS00064; C2 DOMAIN 2; 2.  
DR PROSITE; PSS00479; DAG\_PE\_BIND\_DOM\_1; 1.  
DR PROSITE; PSS00081; DAG\_PE\_BIND\_DOM\_2; 1.  
KW Phorbol-ester binding; Zinc; Repeat; Alternative splicing.  
FT FT DOMAIN 21 96  
FT FT DOMAIN 693 742  
FT FT DOMAIN 802 908  
FT FT DOMAIN 1633 1738  
FT FT VARSPPLIC 1 85  
FT FT VARSPPLIC 1 15  
FT FT VARSPPLIC 1 520  
FT FT VARSPPLIC 521 606

Query Match Best Local Similarity 40.0%; Pred. No. 52;  
Matches 8; Conservative 5; Mismatches 5; Indels 2; Gaps 1;

CY 10 NWAYQVQQ--NKEDAMIEH 27 27.1%; Score 52; DB 1; Length 1813;  
Db 511 NMRYSIOEEDNEKDNWKQH 530

RESULT 7  
GLPK\_SYNY3 STANDARD; PRT; 495 AA.

AC P74260;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Glycerol kinase (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase)  
De (Glycerokinase) (GK).  
OS GLPK OR SLR1672.  
GN Synecocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.  
OX NCBI\_Taxid=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97061201; Pubmed=8905231;  
RA Kaneko T., Sato S., Kojima H., Tanaka A., Asamizu E., Nakamura Y., Miyajima N., Hirosewa M., Sugita M., Sasamoto S., Kimura T., Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Natio K., Okumura S., Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M., Tabata S.;  
RA "Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions." J. DNA Res. 3:109-136(1996).  
RL -1- FUNCTION: KEY ENZYME IN THE REGULATION OF GLYCEROL UPTAKE AND METABOLISM.  
CC -1- CATALYTIC ACTIVITY: ATP + glycerol = ADP + glycerol 3-phosphate.  
CC -1- PATHWAY: Glycerol utilization; rate-limiting step.  
CC -1- SIMILARITY: BELONGS TO THE FUCOKINASE / GLUTONOKINASE / GLYCEROKINASE / XYLULOKINASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D90913; BAA18354.1; -  
 DR HSSP: P08659; IGLC.  
 DR InterPro: IPR000577; FGGY\_kin.  
 DR Pfam: PF00370; FGGY\_1.  
 DR TIGRFA: TIGR01311; glycerol\_kin; 1.  
 DR PROSITE: PS00933; FGGY\_KINASES\_1; FALSE\_NEG.  
 DR PROSITE: PS00445; FGGY\_KINASES\_2; 1; FALSE\_NEG.  
 DR Glycerol metabolism; Transferase; Kinase; ATP-binding;  
 KW Complete proteome.  
 FT NP BIND 156 168 ATP (PROBABLE).  
 FT SEQUENCE 495 AA; 54351 MW; 71AAC71E76816736 CRC64;  
 SQ  
 Query Match 26.6%; Score 51; DB 1; Length 495;  
 Best Local Similarity 34.4%; Pred. No. 17;  
 Matches 11; Conservative 7; Mismatches 10; Indels 4; Gaps 2;  
 Oy 4 YRRKLNMAVQVOONKEDA-WIEHD---VNR 31  
 Db 26 YEGNIVGQATKELTQYPPKAGWEHDALRIK 57  
 RESULT 8  
 ACAD\_ARATH STANDARD; PRT; 1017 AA.  
 ID ACAD\_ARATH  
 AC 09LIX7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Potential calcium-transporting ATPase 13, plasma membrane-type  
 DE (EC 3.6.3.8) (Ca2+-ATPase, isoform 13).  
 GN ACAL3 OR AT3G22910 OR F5N5.8.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=20363099; PubMed=10907853;  
 RA Kaneko T., Katoh T., Sato S., Nakamura A., Asamizu E., Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
 RT Sequence features of the 4,251,695 bp regions covered by 90 Pt, TAC  
 RT and BAC clones.";  
 RL DNA Res. 7:217-221 (2000).  
 CC -1- FUNCTION: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE HYDROLYSIS  
 CC OF ATP COUPLED WITH THE TRANSLLOCATION OF CALCIUM FROM THE CYTOSOL  
 CC OUT OF THE CELL OR INTO ORGANELLES (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + Ca(2+) (in) = ADP + phosphate +  
 CC Ca(2+) (out).  
 CC -1- ENZYME REGULATION: ACTIVATED BY CALMODULIN (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- DOMAIN: THE N-TERMINUS CONTAINS AN AUTOINHIBITORY CALMODULIN-  
 CC BINDING DOMAIN, WHICH BINDS CALMODULIN IN A CALCIUM-DEPENDENT  
 CC FASHION (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY  
 CC (E1-E2 ATPASES). SUBFAMILY IIB.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AP001300; BAB03036.1; -  
 DR HSSP: P04191; IEUL.

DR InterPro: IPR01757; ATPase\_E1-E2.  
 DR InterPro: IPR004014; Cation\_ATPase.  
 DR InterPro: IPR001454; Hlgase/hydlase.  
 DR Pfam: PF00122; E1-E2\_ATPase; 1.  
 DR Pfam: PF00669; Cation\_ATPase\_C; 1.  
 DR Pfam: PF00690; Cation\_ATPase\_N; 1.  
 DR Pfam: PF00702; Hydrolase; 1.  
 DR PRINTS: PR00119; CATALPASE.  
 DR TIGRFA: TIGR01116; Ca\_ATPase; 1.  
 DR PROSITE: PS00154; ATPASE\_E1\_E2; 1.  
 KW Hydrolyase; Calcium transport; Transmembrane; Phosphorylation;  
 KW ATP-binding; Metal-binding; Magnesium; Calmodulin-binding;  
 KW Multigene family; Hypothetical protein.  
 FT DOMAIN 1 147 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 148 168 POTENTIAL.  
 FT DOMAIN 169 186 LUMENAL (POTENTIAL).  
 FT TRANSMEM 187 207 POTENTIAL.  
 FT DOMAIN 208 336 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 337 356 POTENTIAL.  
 FT DOMAIN 357 393 LUMENAL (POTENTIAL).  
 FT TRANSMEM 394 411 POTENTIAL.  
 FT DOMAIN 412 802 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 803 821 POTENTIAL.  
 FT DOMAIN 822 832 LUMENAL (POTENTIAL).  
 FT TRANSMEM 833 853 POTENTIAL.  
 FT DOMAIN 854 873 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 874 896 POTENTIAL.  
 FT DOMAIN 897 905 LUMENAL (POTENTIAL).  
 FT TRANSMEM 906 926 POTENTIAL.  
 FT DOMAIN 927 944 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 945 966 POTENTIAL.  
 FT DOMAIN 967 976 LUMENAL (POTENTIAL).  
 FT TRANSMEM 977 998 POTENTIAL.  
 FT DOMAIN 999 1002 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 20 31 CALMODULIN-BINDING (PROBABLE).  
 FT MOD\_RES 449 449 PHOSPHORYLATION (BY SIMILARITY).  
 FT METAL 747 747 MAGNESIUM (BY SIMILARITY).  
 FT METAL 751 751 MAGNESIUM (BY SIMILARITY).  
 SQ SEQUENCE 1017 AA; 112527 MW; 2F0265CCB862916 CRC64;  
 Query Match 26.6%; Score 51; DB 1; Length 1017;  
 Best Local Similarity 34.6%; Pred. No. 37;  
 Matches 9; Conservative 7; Mismatches 10; Indels 0; Gaps 0;  
 Oy 8 LNMAYQVOONKEDAWIEHDVWRME 33  
 Db 516 ILSWAVELEMEKRYIEHDVWVR 541  
 RESULT 9  
 ALP4\_SCHPO STANDARD; PRT; 784 AA.  
 ID ALP4\_SCHPO  
 AC 09Y705;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Spindle pole body component alp4.  
 GN ALP4 OR SPB365.15.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OS Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OC NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC STRAIN=972;  
 RX MEDLINE=20532503; PubMed=11080156;  
 RA Vardy L., Toda T.;  
 RT "The fission yeast gamma-tubulin complex is required in G(1) phase and  
 RT is a component of the spindle assembly checkpoint.";  
 RL EMBO J. 19:6098-6111 (2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.







```

FT REPEAT 1250 1272 SPECTRIN 5.
FT REPEAT 1287 1342 SPECTRIN 6.
FT REPEAT 1455 1534 SPECTRIN 7.
FT REPEAT 1547 1659 SPECTRIN 8.
FT REPEAT 1815 1891 SPECTRIN 9.
FT REPEAT 1932 2042 SPECTRIN 10.
FT REPEAT 2260 2280 SPECTRIN 11.
FT REPEAT 2372 2395 SPECTRIN 12.
FT REPEAT 2398 2507 SPECTRIN 13.
FT REPEAT 2510 2618 SPECTRIN 14.
FT REPEAT 2621 2728 SPECTRIN 15.
FT REPEAT 2731 2838 SPECTRIN 16.
FT REPEAT 2841 2945 SPECTRIN 17.
FT REPEAT 2987 3024 SPECTRIN 18.
FT REPEAT 3136 3163 SPECTRIN 19.
FT REPEAT 3187 3274 SPECTRIN 20.
FT REPEAT 3277 3383 SPECTRIN 21.
FT REPEAT 3386 3492 SPECTRIN 22.
FT REPEAT 3495 3601 SPECTRIN 23.
FT REPEAT 3604 3673 SPECTRIN 24.
FT REPEAT 3713 3819 SPECTRIN 25.
FT REPEAT 3832 3927 SPECTRIN 26.
FT REPEAT 3982 4043 SPECTRIN 27.
FT REPEAT 4046 4152 SPECTRIN 28.
FT REPEAT 4155 4262 SPECTRIN 29.
FT REPEAT 4265 4371 SPECTRIN 30.
FT REPEAT 4374 4481 SPECTRIN 31.
FT REPEAT 4484 4590 SPECTRIN 32.
FT REPEAT 4593 4700 SPECTRIN 33.
FT REPEAT 4707 4808 SPECTRIN 34.
FT REPEAT 4811 4917 SPECTRIN 35.
FT REPEAT 4920 4985 SPECTRIN 36.
FT REPEAT 5030 5054 SPECTRIN 37.
FT CA_BIND 5096 5107 EF-HAND 1 (POTENTIAL).
FT 5132 5143 EF-HAND 2 (POTENTIAL).
FT 5276 5283 POLY-SER.
FT DOMAIN 5370 5370 4 X 4 AA. TANDEM REPEATS OF [GS]-S-R-(AR).
FT CONFLECT 1 MSSSDELTLSERCSRCSRYSRSGSLSPCPD
TLPWNLPIHEOKRKSDSLDPAERAVVY -> MPPVLM
AGIPGRVGSGLPPLPGFKQFCSASRVAVI (IN REF.
2).
VAISSSEDEGNLRFVYELLS -> GRSAPLKMALISDLGN
YCL (IN REF. 1).
A -> T (IN REF. 2 AND 3).
V -> A (IN REF. 1).
E -> D (IN REF. 1).
E -> K (IN REF. 2).
M -> V (IN REF. 2).
C -> Y (IN REF. 1).
MISSING (IN REF. 2).
SILPSVG -> EYRLFTI (IN REF. 3).
O -> R (IN REF. 2 AND 4).
S -> T (IN REF. 2).
MISSING (IN REF. 2).
SEQUENCE 5430 AA; 620346 MW; 91ADBYF7580B440B CRC64;

```

```

Query Match 26.8%; Score 50; DB 1; Length 5430;
Best Local Similarity 40.7%; Pred. No. 3.3e+02;
Matches 11; Conservative 6; Mismatches 10; Indels 0; Gaps 0;
QY 5 RYKLLNMAVOQONKEDAWIEHWVR 31
Db 5048 RQRLNDALDRLEIKFANPDPVWR 5074

```

```

RESULT 12
YQ04_BACAN STANDARD; PRT; 95 AA.
AC G9RNZ8;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein PX02-04.

```

```

GN PX02-04.
OS Bacillus anthracis.
OG Plasmid PX02.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RA Okinaka R.T., Cloud K., Hampton O., Hill K.K., Keim P., Lamke G.,
RA Kumano S., Mautner D., Martinez Y., Svensson R., Tatam L.R.,
RA Brown A.E., Jackson P.J.;
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF188935; AAF13609.1; -.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 95 AA; 11613 MW; 2C622DD9CFA866F CRC64;

```

```

Query Match 25.8%; Score 49.5; DB 1; Length 95;
Best Local Similarity 35.5%; Pred. No. 4.2;
Matches 11; Conservative 7; Mismatches 8; Indels 5; Gaps 1;
QY 1 RRSRYKLLNMAV-----OOVQONKEDAWIE 26
Db 29 RRYRWBYIMWFQHFQNFQYSRNKKEDFDE 59

```

```

RESULT 13
DAPD_BUCAI STANDARD; PRT; 274 AA.
AC P57353;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase
(EC 2.3.1.117) (tetrahydronicotinamide N-succinyltransferase)
DE (TNP succinyltransferase) (tetrahydropicolinate succinylase).
DE DAPD OR BU229.
GN Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS Buchnera aphidicola)
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Wainabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS."
RT Nature 407:81-86(2000).
RL
CC -1- CATALYTIC ACTIVITY: Succinyl-CoA + 2,3,4,5-tetrahydropyridine-2-
CC carboxylate = CoA + N-succinyl-L-2-amino-6-oxoheptanoate.
CC -1- PATHWAY: POUFUR STEP IN THE BIOSYNTHESIS OF DIMINOPIMELATE AND
CC LYSINE FROM ASPARTATE SEMIALDEHYDE.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE CYSE/LACA/LPXA/NODL FAMILY OF
CC ACETYLTRANSFERASES. COMPOSED OF MULTIPLE REPEATS OF [LIV]-G-X(4).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AP001118; BAB12944.1; -.

```



SQ SEQUENCE 940 AA; 107111 MM; 28054683FB9D0404 CRC64;

Query Match 25.8%; Score 49.5; DB 1; Length 940;

Best Local Similarity 31.6%; Pred. No. 54;

Matches 12; Conservative 8; Mismatches 9; Indels 9; Gaps 2;

QY 4 YR-YKLIW-----AYQVQONKEDAMIEHDVWRM 32

DB 172 YRGYLVNWDVPLQATALADDEVEYERKDGMLYIRYRM 209

Search completed: March 26, 2003, 16:45:07

Job time : 11.8837 secs

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: March 26, 2003, 16:44:15 ; Search time 19.3721 Seconds  
(without alignments)  
168.726 Million cell updates/sec

Title: US-10-010-667a-20

Perfect score: 192  
Sequence: 1 RRSYRYKLIMAWYQVQNKEDAMIEHDVWMEI 34

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	62.5	32.6	480	2 B70446	hypothetical prote
2	58.5	30.5	1418	2 S40764	hypothetical prote
3	58	30.2	900	2 E69631	galactosamine-cont
4	56	29.2	150	2 T28409	ORF MSV248 probabl
5	55.5	28.9	502	2 S49113	hypothetical prote
6	53	27.6	152	2 C69546	hypothetical prote
7	52	27.1	79	2 E69866	hypothetical prote
8	52	27.1	1734	2 A41101	phorbol ester-bind
9	52	27.1	1813	2 T19295	hypothetical prote
10	51.5	26.8	713	2 G64606	hydatonin utilizat
11	51	26.6	90	2 AB2102	hypothetical prote
12	51	26.6	444	2 T27222	hypothetical prote
13	51	26.6	495	1 S75895	probable L-xyulok
14	51	26.6	793	2 C83260	hypothetical prote
15	50.5	26.3	201	2 AG2434	hypothetical prote
16	50.5	26.3	402	2 AG1839	hypothetical prote
17	50.5	26.3	712	2 G71907	probable hydatonin
18	50	26.0	204	2 AC2411	hypothetical prote
19	50	26.0	784	2 T43510	probable spinole p
20	50	26.0	816	2 T00919	hypothetical prote
21	50	26.0	1575	2 S68448	synaptojanin, 170K
22	50	26.0	1697	2 T00079	hypothetical prote
23	50	26.0	1908	2 T42707	hypothetical prote
24	49.5	25.8	133	2 E84502	hypothetical prote
25	49.5	25.8	224	2 A10697	probable pathogeni
26	49.5	25.8	274	2 H84956	2,3,4,5-tetrahydro
27	49.5	25.8	350	2 H64458	hypothetical prote
28	49.5	25.8	414	2 T16602	hypothetical prote
29	49.5	25.8	787	2 E71537	probable 60kda inn

30	49.5	25.8	940	2 P86502	alyl tRNA synthet
31	49.5	25.8	940	2 B72120	valine-tRNA ligase
32	49.5	25.8	1391	2 T20642	hypothetical prote
33	49.5	25.8	1397	2 E87998	protein F09C3.1 (f
34	49	25.5	257	2 A82691	tRNA pseudouridine
35	49	25.5	263	2 A82069	hypothetical prote
36	49	25.5	383	2 H84700	hypothetical prote
37	49	25.5	388	2 T40609	hypothetical prote
38	49	25.5	500	2 AE2032	glycerol kinase (f
39	49	25.5	1120	2 H88449	protein F54D8.1 (f
40	49	25.5	2925	2 T00133	RNA-directed RNA p
41	48.5	25.3	266	2 C90257	translation initia
42	48.5	25.3	304	2 T04751	hypothetical prote
43	48.5	25.3	339	2 T33477	hypothetical prote
44	48.5	25.3	505	2 B81021	conserved hypothet
45	48.5	25.3	638	2 T20944	hypothetical prote

## ALIGNMENTS

RESULT 1  
B70446  
hypothetical protein aq\_1687 - Aquifex aeolicus  
C/Species: Aquifex aeolicus  
C/Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999  
C/Accession: B70446  
R/Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove  
V.  
Nature 392, 353-358, 1998  
A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A/Reference number: A70300; MUID:98196666; PMID:9537320  
A/Accession: B70446  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-480 <AGP>  
A/Cross-references: GB:AE000751; NID:g2984013; PIDN:AAC07560.1; PID:g2984016; GB:AE00065  
A/Experimental source: strain VFS  
C/Genetics:  
A/Gene: aq\_1687

Query Match 32.6%; Score 62.5; DB 2; Length 480;  
Best Local Similarity 25.0%; Pred. No. 1.8;  
Matches 12; Conservative 17; Mismatches 4; Indels 15; Gaps 2;

QY 1 RRSYR-----YKLIMAWYQVQNKEDAMIEHDVWMEI 33  
DB 10 RKTYYRLKIPAKVSDGKRYVLDMSYEGFRLEKSKEDVFEKDKYKVK 57

RESULT 2  
S40764  
hypothetical protein ZK512.3 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 24-Nov-1999  
C/Accession: S40764  
R/Hawkins, T.; Ainscough, R.  
submitted to the EMBL Data Library, February 1993  
A/Reference number: S40759  
A/Accession: S40764  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1418 <HAW>  
A/Cross-references: EMBL:Z22177; NID:g297989; PID:g297992  
C/Genetics:  
A/Insertions: 199/1; 238/1; 290/2; 529/3; 557/3; 588/3; 677/2; 733/3; 772/3; 846/3; 946/1; 1  
C/Superfamily: Caenorhabditis elegans hypothetical protein ZK512.3

Query Match 30.5%; Score 58.5; DB 2; Length 1418;  
Best Local Similarity 43.5%; Pred. No. 19;  
Matches 10; Conservative 5; Mismatches 5; Indels 3; Gaps 1;

QY 15 QVQNKED---AWIEHDVWMEI 34

Db 1355 EVANENDTDNMTAEHDVMAISL 1377

## RESULT 3

E69631 galactosamine-containing minor teichoic acid biosynthesis ggaB - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999

C:Accession: E69631

R:Kunitz, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Beret, C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho, A.; Erlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardinois, A.; Laubert, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel, Y., M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot, A.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: E69631

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-900 <KUN>

A:Cross-references: GB:Z99122; GB:AL009126; NID:92636029; PIDN:CA81585.1; PID:el184474;

A:Experimental source: strain 168

A:Genetics:

A:Gene: ggaB

Query Match 30.2%; Score 58; DB 2; Length 900;

Best Local Similarity 33.3%; Pred. No. 14;

Matches 11; Conservative 8; Mismatches 14; Indels 0; Gaps 0;

Db 591 RSRKILTLSSKVTSTADLVNVEFFNMEI 623

Query 2 RSYRYKLWMAVOQONKEDAMIEHDVMEI 34

Db 591 RSRKILTLSSKVTSTADLVNVEFFNMEI 623

RESULT 4

T28409 ORF MSV248 Probable inhibitor of apoptosis protein (IAP) - Melanoplus sanguinipes entomoc

C:Species: Melanoplus sanguinipes entomopoxvirus

C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000

C:Accession: T28409

R:Afonso, C.L.; Tulman, B.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.

J. Virol. 73, 533-552, 1999

A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.

A:Reference number: Z20484; MUID:99102612; PMID:9847359

A:Accession: T28409

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Cross-references: EMBL:AF063866; NID:94049647; PIDN:AAC97724.1; PID:94049764

hypothetical protein 2 - Microcystis aeruginosa

C:Species: Microcystis aeruginosa

C:Date: 01-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 23-Mar-2001

C:Accession: S49113

R:Juerchott, K.; Boerner, T.

submitted to the EMBL Data Library, November 1993

A:Description: Sequence of the cyanobacterial plasmid pMA1 from Microcystis aeruginosa H

A:Reference number: S49112

A:Accession: S49113

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-502 <JUR>

A:Cross-references: EMBL:Z28337; NID:9509352; PIDN:CA82191.1; PID:9509354

C:Superfamily: Microcystis aeruginosa hypothetical protein 2

Query Match 28.9%; Score 55.5; DB 2; Length 502;

Best Local Similarity 25.0%; Pred. No. 16;

Matches 12; Conservative 7; Mismatches 14; Indels 15; Gaps 1;

Db 392 QAVREKLWMAVRVYDGPQLPEEFKALQAEANKQWENNMSWDEL 439

Query 2 RSYRYKLWMA-----YQOVQONKEDAMIEHDVMEI 34

Db 392 QAVREKLWMAVRVYDGPQLPEEFKALQAEANKQWENNMSWDEL 439

RESULT 6

C69546 hypothetical protein AF2371 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999

C:Accession: C69546

R:Klein, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Uitterback, T.; Cotton, M.D.; Spriggs, T.; Arriach, P.; Kaine, B.P.; Sykes, S.B.

Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeor

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: C69546

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-152 <KLE>

A:Cross-references: GB:AE001112; GB:AE000782; NID:92689435; PIDN:AAB91236.1; PID:9265072;

Query Match 42.6%; Score 53; DB 2; Length 152;

Best Local Similarity 27.9%; Pred. No. 9.4;

Matches 9; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Db 21 RYKLSMLVDRYERNRVGAYI 41

Query 5 RYKLSMLVDRYERNRVGAYI 25

Db 21 RYKLSMLVDRYERNRVGAYI 41

RESULT 7

E69966 hypothetical protein yqkC - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999

C:Accession: E69966

R:Kunitz, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Beret, C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho, A.; Erlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardinois, A.; Laubert, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel, Y., M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot, A.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeor

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: C69546

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-152 <KLE>

A:Cross-references: GB:AE001112; GB:AE000782; NID:92689435; PIDN:AAB91236.1; PID:9265072;

Query Match 42.6%; Score 53; DB 2; Length 152;

Best Local Similarity 27.9%; Pred. No. 9.4;

Matches 9; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Db 21 RYKLSMLVDRYERNRVGAYI 41

Query 5 RYKLSMLVDRYERNRVGAYI 25

Db 21 RYKLSMLVDRYERNRVGAYI 41

A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A/Reference number: A69580; MUID:9804033; PMID:9384377  
 A/Accession: B69966  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-79 <NON>  
 A/Cross-references: GB:299116; GB:AL009126; NID:G2634723; PIDN:CAB14297.1; PID:el185634;  
 A/Experimental source: strain 168  
 C/Genetics:  
 A/Gene: yqkC

Query Match 27.1%; Score 52; DB 2; Length 79;  
 Best Local Similarity 28.1%; Pred. No. 6.3;  
 Matches 9; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

QY 2 RSRRYKLMMAYQOVQONKEDAMIEHDVWRME 33  
 Db 23 QNHPTLHMSTIGASIKDVLQDEMTFE 54

RESULT 8  
 A41101  
 photol ester-binding protein unc-13 - *Caenorhabditis elegans*  
 C/Species: *Caenorhabditis elegans*  
 C/Date: 20-Mar-1992 #sequence\_revision 20-Mar-1992 #text\_change 23-Feb-1997  
 C/Accession: A41101  
 R/Martyna, I.N.; Brenner, S.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 5729-5733, 1991  
 A>Title: A photol ester/diacylglycerol-binding protein encoded by the unc-13 gene of *Caenorhabditis elegans*  
 A/Reference number: A41101; MUID:91288538; PMID:2062851  
 A/Accession: A41101  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-1734 <MAR>  
 A/Cross-references: GB:M62830  
 C/Superfamily: protein kinase C zinc-binding repeat homology  
 C/Keywords: phosphoprotein  
 F:615-664/Domain: protein kinase C zinc-binding repeat homology <K2>

Query Match 27.1%; Score 52; DB 2; Length 1734;  
 Best Local Similarity 40.0%; Pred. No. 1.7e+02;  
 Matches 8; Conservative 5; Mismatches 5; Indels 2; Gaps 1;

QY 10 NMAYQOVQO--NKEDAMIEH 27  
 Db 430 NMRYSIOEDNEKDMKQH 449

RESULT 9  
 T19295  
 hypothetical protein ZK524.2a - *Caenorhabditis elegans*  
 C/Species: *Caenorhabditis elegans*  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 02-Sep-2000  
 C/Accession: T19295; T19311; T27889  
 R/Gardner, A.  
 submitted to the EMBL Data Library, September 1996  
 A/Reference number: Z19103  
 A/Accession: T19295  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-1813 <MIL>  
 A/Cross-references: EMBL:Z79694; PIDN:CAB01966.1; GSPDB:GN00019; CESP:ZK524.2a  
 A/Experimental source: clone C15M11  
 R/Lloyd, C.  
 submitted to the EMBL Data Library, March 1997  
 A/Reference number: Z19198  
 A/Accession: T19311  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-1813 <MIL>  
 A/Cross-references: EMBL:Z92779; PIDN:CAB01713.1; GSPDB:GN00019; CESP:ZK524.2a  
 A/Experimental source: clone C44E1  
 R/Gardner, A.

submitted to the EMBL Data Library, June 1996  
 A/Reference number: Z20435  
 A/Accession: T27889  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-1813 <MIL>  
 A/Cross-references: EMBL:Z73912; PIDN:CAA98147.1; GSPDB:GN00019; CESP:ZK524.2a  
 A/Experimental source: clone ZK524  
 C/Genetics:  
 A/Gene: CESP:ZK524.2a  
 A/Map position: 1  
 A/Introns: 25/1; 35/1; 68/2; 105/2; 149/1; 170/3; 186/1; 235/3; 412/2; 434/2; 500/2; 540/  
 1748/1; 1781/3  
 C/Superfamily: protein kinase C zinc-binding repeat homology  
 F:693-742/Domain: protein kinase C zinc-binding repeat homology <K2N>

Query Match 27.1%; Score 52; DB 2; Length 1813;  
 Best Local Similarity 40.0%; Pred. No. 1.8e+02;  
 Matches 8; Conservative 5; Mismatches 5; Indels 2; Gaps 1;

QY 10 NMAYQOVQO--NKEDAMIEH 27  
 Db 511 NMRYSIOEDNEKDMKQH 530

RESULT 10  
 G64606  
 hydantoin utilization protein A - *Helicobacter pylori* (strain 26695)  
 C/Species: *Helicobacter pylori*  
 C/Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 16-Feb-2001  
 C/Accession: G64606  
 R/Tomb, J.F.; White, O.; Kariavase, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;  
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.; McKenney  
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weldman, J.W.; Fujii, C.; Bowman, C.; Watthey, L.  
 Nature 388, 539-547, 1997  
 A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Frazer, C.A.  
 A>Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.  
 A/Reference number: A64520; MUID:97394467; PMID:9252185  
 A/Accession: G64606  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-713 <TOM>  
 A/Cross-references: GB:AE00582; GB:AE00511; NID:G2313812; PIDN:AA00746.1; PID:G231381  
 C/Superfamily: Pseudomonas D-amino acid hydantoin hydrolase (ATP-hydrolyzing) hyuA

Query Match 26.8%; Score 51.5; DB 2; Length 713;  
 Best Local Similarity 30.3%; Pred. No. 76;  
 Matches 10; Conservative 5; Mismatches 3; Indels 15; Gaps 1;

QY 1 RSRRYKLMMAYQOVQONKEDAMIEHDVWRME 33  
 Db 653 RRYRHK-----KMDADYQOME 670

RESULT 11  
 AB2102  
 hypothetical protein aer2369 [imported] - *Nostoc* sp. (strain PCC 7120)  
 C/Species: *Nostoc* sp.  
 A/Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120  
 C/Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C/Accession: AB2102  
 R/Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuchi, S.  
 DNA Res. 8, 205-213, 2001  
 A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anat*  
 A/Reference number: AB1807; MUID:21595285; PMID:11759840  
 A/Accession: AB2102  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-90 <KIR>  
 A/Cross-references: GB:BA000019; PIDN:BA074068.1; PID:G17131461; GSPDB:GN00179  
 A/Experimental source: strain PCC 7120  
 C/Genetics:





GenCore version 5.1.4\_p5 4578  
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 16:49:57 ; Search time 22.9302 Seconds  
(without alignments)  
87.073 Million cell updates/sec

Title: US-10-010-667a-20

Perfect score: 192  
Sequence: 1 RRSRYKLIMAYQQVQONKEDAMIEHDVWMEI 34

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 237916 seqs, 58723674 residues

Total number of hits satisfying chosen parameters: 237916

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications\_AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	192	100.0	34	9 US-10-011-095-20	Sequence 20, Appl
2	192	100.0	34	9 US-10-010-667a-20	Sequence 20, Appl
3	192	100.0	267	10 US-09-747-835A-50	Sequence 51, Appl
4	192	100.0	267	10 US-09-747-835A-51	Sequence 51, Appl
5	192	100.0	339	9 US-10-012-896-879	Sequence 879, Appl
6	192	100.0	339	9 US-09-802-520-11	Sequence 11, Appl
7	192	100.0	339	9 US-09-895-793-879	Sequence 879, Appl
8	192	100.0	339	9 US-09-895-814-879	Sequence 879, Appl
9	192	100.0	339	9 US-10-011-095-2	Sequence 2, Appl
10	192	100.0	339	9 US-10-010-667a-2	Sequence 2, Appl
11	192	100.0	339	10 US-09-759-143-879	Sequence 879, Appl
12	192	100.0	339	10 US-09-780-663-879	Sequence 879, Appl
13	192	100.0	339	10 US-09-822-827-879	Sequence 879, Appl
14	192	100.0	141	10 US-09-963-896-1	Sequence 1, Appl
15	192	100.0	141	10 US-10-011-095-8	Sequence 8, Appl
16	192	100.0	173	9 US-10-010-667a-8	Sequence 8, Appl
17	192	100.0	237	10 US-09-747-835A-15	Sequence 15, Appl
18	192	100.0	488	10 US-09-747-835A-13	Sequence 13, Appl
19	192	100.0	490	9 US-09-802-520-1	Sequence 1, Appl

20	85	44.3	15	9 US-10-011-095-32	Sequence 32, Appl
21	85	44.3	15	9 US-10-010-667a-32	Sequence 32, Appl
22	52	27.1	38	10 US-09-864-761-38670	Sequence 38670, A
23	52	27.1	1114	10 US-09-740-046-10	Sequence 10, Appl
24	51.5	26.8	491	10 US-09-881-752A-128	Sequence 128, Appl
25	51	26.6	487	9 US-09-738-626-3647	Sequence 3647, Ap
26	50	26.0	370	9 US-09-809-391-436	Sequence 436, Appl
27	50	26.0	469	10 US-09-925-300-1494	Sequence 1494, Ap
28	50	26.0	1193	12 US-10-040-919-2	Sequence 2, Appl
29	48.5	25.3	151	9 US-09-860-670-145	Sequence 145, Appl
30	48.5	25.3	151	10 US-09-764-869-974	Sequence 974, Appl
31	48.5	25.3	654	10 US-09-963-528-10	Sequence 10, Appl
32	48	25.0	496	9 US-09-738-626-5483	Sequence 5483, Ap
33	48	25.0	825	9 US-10-101-467A-922	Sequence 922, Appl
34	47.5	24.7	4150	9 US-09-808-880-2	Sequence 2, Appl
35	47	24.5	274	10 US-09-739-907-171	Sequence 171, Appl
36	47	24.5	441	10 US-09-991-258-16	Sequence 16, Appl
37	47	24.5	525	10 US-09-739-907-79	Sequence 79, Appl
38	47	24.5	669	9 US-10-083-357-1336	Sequence 1336, Ap
39	47	24.5	823	9 US-10-101-467A-923	Sequence 923, Appl
40	47	24.5	1503	9 US-10-007-706-1	Sequence 1, Appl
41	46.5	24.2	135	10 US-09-764-864-1191	Sequence 1191, Ap
42	46.5	24.2	215	9 US-09-978-295A-150	Sequence 150, Appl
43	46.5	24.2	215	9 US-09-978-697-150	Sequence 150, Appl
44	46.5	24.2	215	9 US-09-978-192A-150	Sequence 150, Appl
45	46.5	24.2	215	9 US-09-999-832A-150	Sequence 150, Appl

#### ALIGNMENTS

RESULT 1  
US-10-011-095-20  
; Sequence 20, Application US/10011095  
; Publication No. US20030045682A1  
; GENERAL INFORMATION:  
; APPLICANT: Afar, Daniel  
; APPLICANT: Hubert, Rene S.  
; APPLICANT: Leong, Kahan  
; APPLICANT: Raftano, Douglas B.  
; APPLICANT: Safetan, Douglas C.  
; APPLICANT: Mitchell, Steve Chappell  
; TITLE OF INVENTION: ANTIBODIES IMMUNOSPECIFIC FOR STRAP1 (AS AMENDED)  
; FILE REFERENCE: 511582001610  
; CURRENT APPLICATION NUMBER: US/10/011,095  
; PRIOR FILING DATE: 2001-12-06  
; PRIOR APPLICATION NUMBER: 09/323,873  
; PRIOR FILING DATE: 1999-06-01  
; PRIOR APPLICATION NUMBER: 60/087,520  
; PRIOR FILING DATE: 1998-06-01  
; PRIOR APPLICATION NUMBER: 60/091,183  
; PRIOR FILING DATE: 1998-06-30  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 20  
; LENGTH: 34  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: STRAP-1 peptide  
US-10-011-095-20

Query Match 100.0% Score 192; DB 9; Length 34;  
Best Local Similarity 100.0%; Pred. No. 1.1e-19;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRSRYKLIMAYQQVQONKEDAMIEHDVWMEI 34  
DB 1 RRSRYKLIMAYQQVQONKEDAMIEHDVWMEI 34  
RESULT 2  
US-10-010-667a-20

Sequence 20, Application US/10010667A  
Publication No. US20030055217A1  
GENERAL INFORMATION:  
APPLICANT: Afar, Daniel  
APPLICANT: Hubert, Rene S.  
APPLICANT: Leong, Kahen  
APPLICANT: Raitano, Arthur B.  
APPLICANT: Saffran, Douglas C.  
APPLICANT: Mitchell, Steve Chappell  
TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS  
FILE REFERENCE: 51582001601  
CURRENT APPLICATION NUMBER: US/10/010,667A  
PRIOR FILING DATE: 2001-12-06  
PRIOR APPLICATION NUMBER: 09/323,873  
PRIOR FILING DATE: 1999-06-01  
PRIOR APPLICATION NUMBER: 60/087,520  
PRIOR FILING DATE: 1998-06-01  
PRIOR APPLICATION NUMBER: 60/091,183  
PRIOR FILING DATE: 1998-06-30  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: PasteSeq for Windows Version 4.0  
SEQ ID NO 20  
LENGTH: 34  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: STEAP-1 peptide  
US-10-010-667A-20

Query Match 100.0%; Score 192; DB 9; Length 34;  
Best Local Similarity 100.0%; Pred. No. 1,1e-19;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RRSRYKLNMAYQVOQNKEDAMIEHDVWMEI 34  
Db 1 RRSRYKLNMAYQVOQNKEDAMIEHDVWMEI 34

RESULT 3  
US-09-747-835A-50  
Sequence 50, Application US/09747835A  
GENERAL INFORMATION:  
APPLICANT: Yamazaki, Victoria  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Zhou, Ping  
APPLICANT: Wang, Dunrui  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Asundi, Vinod  
APPLICANT: Dmanac, Radote T  
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-LIKE  
FILE REFERENCE: HYS-37CIP  
CURRENT APPLICATION NUMBER: US/09/747,835A  
PRIOR FILING DATE: 2002-03-08  
PRIOR APPLICATION NUMBER: US 09/729,739  
PRIOR FILING DATE: 2000-12-04  
PRIOR APPLICATION NUMBER: US 09/653,450  
PRIOR FILING DATE: 2000-08-31  
PRIOR APPLICATION NUMBER: US 09/620,312  
PRIOR FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: US 09/598,042  
PRIOR FILING DATE: 2000-06-20  
PRIOR APPLICATION NUMBER: US 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: US 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 50

LENGTH: 267  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-747-835A-50

Query Match 100.0%; Score 192; DB 10; Length 267;  
Best Local Similarity 100.0%; Pred. No. 1e-18;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RRSRYKLNMAYQVOQNKEDAMIEHDVWMEI 34  
Db 119 RRSRYKLNMAYQVOQNKEDAMIEHDVWMEI 152

RESULT 4  
US-09-747-835A-51  
Sequence 51, Application US/09747835A  
GENERAL INFORMATION:  
APPLICANT: Yamazaki, Victoria  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Zhou, Ping  
APPLICANT: Wang, Dunrui  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Asundi, Vinod  
APPLICANT: Dmanac, Radote T  
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-LIKE  
FILE REFERENCE: HYS-37CIP  
CURRENT APPLICATION NUMBER: US/09/747,835A  
PRIOR FILING DATE: 2002-03-08  
PRIOR APPLICATION NUMBER: US 09/729,739  
PRIOR FILING DATE: 2000-12-04  
PRIOR APPLICATION NUMBER: US 09/653,450  
PRIOR FILING DATE: 2000-08-31  
PRIOR APPLICATION NUMBER: US 09/620,312  
PRIOR FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: US 09/598,042  
PRIOR FILING DATE: 2000-06-20  
PRIOR APPLICATION NUMBER: US 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: US 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 51  
LENGTH: 267  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-747-835A-51

Query Match 100.0%; Score 192; DB 10; Length 267;  
Best Local Similarity 100.0%; Pred. No. 1e-18;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RRSRYKLNMAYQVOQNKEDAMIEHDVWMEI 34  
Db 119 RRSRYKLNMAYQVOQNKEDAMIEHDVWMEI 152

RESULT 5  
US-10-012-896-879  
Sequence 879, Application US/10012896  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Dillon, Davin C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Harlocker, Susan L.  
APPLICANT: Jiang, Yuguo  
APPLICANT: Kalos, Michael D.

```

; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Mantnabe, Yoshihiro
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012,896
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 879
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-896-879

Query Match          100.0%; Score 192; DB 9; Length 339;
Best Local Similarity 100.0%; Pred. No. 1.3e-18;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

Qy 1 RRSRYKLLNMAVQVQONKEDAMIEHDVWRMEI 34
Db 185 RRSRYKLLNMAVQVQONKEDAMIEHDVWRMEI 218

RESULT 6
US-09-802-520-11
; Sequence 11, Application US/09802520
; Publication No. US20020187472A1
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Paris, Mary
; APPLICANT: Chen, Hui-Mei
; APPLICANT: Ison, Craig H.
; TITLE OF INVENTION: STEAP-RELATED PROTEIN
; FILE REFERENCE: PC-0037 US
; CURRENT APPLICATION NUMBER: US/09/802,520
; CURRENT FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PERL Program
; SEQ ID NO 11
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020187472A1 96572948
US-09-802-520-11

Query Match          100.0%; Score 192; DB 9; Length 339;
Best Local Similarity 100.0%; Pred. No. 1.3e-18;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

Qy 1 RRSRYKLLNMAVQVQONKEDAMIEHDVWRMEI 34
Db 185 RRSRYKLLNMAVQVQONKEDAMIEHDVWRMEI 218
```

```

RESULT 7
US-09-895-793-879
; Sequence 879, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895,793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 879
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-793-879

Query Match          100.0%; Score 192; DB 9; Length 339;
Best Local Similarity 100.0%; Pred. No. 1.3e-18;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

Qy 1 RRSRYKLLNMAVQVQONKEDAMIEHDVWRMEI 34
Db 185 RRSRYKLLNMAVQVQONKEDAMIEHDVWRMEI 218

RESULT 8
US-09-895-814-879
; Sequence 879, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
```

```

; APPLICANT: Vinals de Bascois, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fangier, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 879
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-895-814-879

Query Match          100.0%; Score 192; DB 9; Length 339;
Best Local Similarity 100.0%; Pred. No. 1.3e-18;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRSRYKLLMAYQVOONKEDAMIEHDVWRMEI 34
DB 185 RRSRYKLLMAYQVOONKEDAMIEHDVWRMEI 218

RESULT 9
US-10-011-095-2
; Sequence 2, Application US/10011095
; Publication No. US20030045682A1
; GENERAL INFORMATION:
; APPLICANT: Afari, Daniel
; APPLICANT: Hubert, Rene S.
; APPLICANT: Leong, Kahen
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Saffran, Douglas C.
; APPLICANT: Mitchell, Steve Chappell
; TITLE OF INVENTION: ANTIBODIES IMMUNOSPECIFIC FOR STEAP1 (AS AMENDED)
; FILE REFERENCE: 51182001610
; CURRENT APPLICATION NUMBER: US/10/011,095
; CURRENT FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: 09/323,873
; PRIOR FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-01
; PRIOR APPLICATION NUMBER: 60/091,183
; PRIOR FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DNA
; US-10-011-095-2

Query Match          100.0%; Score 192; DB 9; Length 339;
Best Local Similarity 100.0%; Pred. No. 1.3e-18;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRSRYKLLMAYQVOONKEDAMIEHDVWRMEI 34
DB 185 RRSRYKLLMAYQVOONKEDAMIEHDVWRMEI 218

RESULT 10
US-10-010-667A-2
; Sequence 2, Application US/10010667A
; Publication No. US20030055217A1
; GENERAL INFORMATION:
; APPLICANT: Afari, Daniel
; APPLICANT: Hubert, Rene S.
; APPLICANT: Leong, Kahen
```

```

; APPLICANT: Raitano, Arthur B.
; APPLICANT: Saffran, Douglas C.
; APPLICANT: Mitchell, Steve Chappell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; FILE REFERENCE: 51182001601
; CURRENT APPLICATION NUMBER: US/10/010,667A
; CURRENT FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: 09/323,873
; PRIOR FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-01
; PRIOR APPLICATION NUMBER: 60/091,183
; PRIOR FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-010-667A-2

Query Match          100.0%; Score 192; DB 9; Length 339;
Best Local Similarity 100.0%; Pred. No. 1.3e-18;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRSRYKLLMAYQVOONKEDAMIEHDVWRMEI 34
DB 185 RRSRYKLLMAYQVOONKEDAMIEHDVWRMEI 218

RESULT 11
US-09-759-143-879
; Sequence 879, Application US/09759143
; Patent No. US20020022248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqun
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fangier, Gary R.
; APPLICANT: Retler, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yassir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759,143
; CURRENT FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 879
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-759-143-879

Query Match          100.0%; Score 192; DB 10; Length 339;
Best Local Similarity 100.0%; Pred. No. 1.3e-18;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRSRYKLLMAYQVOONKEDAMIEHDVWRMEI 34
DB 185 RRSRYKLLMAYQVOONKEDAMIEHDVWRMEI 218
```



FILE REFERENCE: 51582001610  
CURRENT APPLICATION NUMBER: US/10/011,095  
CURRENT FILING DATE: 2001-12-06  
PRIOR APPLICATION NUMBER: 09/323,873  
PRIOR FILING DATE: 1999-06-01  
PRIOR APPLICATION NUMBER: 60/087,520  
PRIOR FILING DATE: 1998-06-01  
PRIOR APPLICATION NUMBER: 60/091,183  
PRIOR FILING DATE: 1998-06-30  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 8  
LENGTH: 173  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-011-095-8

Query Match 55.7%; Score 107; DB 9; Length 173;  
Best Local Similarity 58.8%; Pred. No. 2.3e-07;  
Matches 20; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 RRSYRYKLNMAYQVQONKEDAVIEHDVWPMET 34  
DB 81 RRSERYLPLNMAYQVAVANTENSNHEEVRRIEM 114

Search completed: March 26, 2003, 17:04:16  
Job time: 23.9302 secs

GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 16:44:40 ; Search time 21.7442 Seconds  
(without alignments)  
46.007 Million cell updates/sec

Title: US-10-010-667a-20

Perfect score: 192  
Sequence: 1 RRSRYRYKLIMWAVQVQONKEDAWIEHDVWRMEI 34

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	192	100.0	34	US-09-323-873A-20	Sequence 20, Appl
2	192	100.0	339	US-09-323-873A-2	Sequence 2, Appl
3	107	55.7	141	US-09-083-521-1	Sequence 1, Appl
4	107	55.7	173	US-09-323-873A-8	Sequence 8, Appl
5	85	44.3	15	US-09-323-873A-32	Sequence 32, Appl
6	50	26.0	370	US-09-149-676-436	Sequence 436, Appl
7	48.5	25.3	654	US-08-560-005-10	Sequence 10, Appl
8	48.5	25.3	654	US-09-418-540-10	Sequence 10, Appl
9	47.5	24.7	178	US-09-134-001C-3412	Sequence 3412, Ap
10	47.5	24.7	4150	US-09-428-517-2	Sequence 2, Appl
11	46	24.0	236	US-09-121-879-4	Sequence 4, Appl
12	46	24.0	236	US-09-332-319-4	Sequence 4, Appl
13	46	24.0	236	US-09-239-867-2	Sequence 2, Appl
14	45	23.4	371	US-08-746-789A-2	Sequence 2, Appl
15	45	23.4	885	US-09-342-648-9	Sequence 9, Appl
16	45	23.4	1417	US-08-559-303B-78	Sequence 78, Appl
17	45	23.4	1417	US-08-781-891-78	Sequence 78, Appl
18	45	23.4	1417	US-09-175-828-78	Sequence 78, Appl
19	44	22.9	69	US-09-134-001C-3885	Sequence 3885, Ap
20	44	22.9	303	US-09-372-934-2	Sequence 1, Appl
21	44	22.9	307	US-08-713-828-1	Sequence 1, Appl
22	44	22.9	307	US-08-919-627-1	Sequence 1, Appl
23	44	22.9	307	US-09-096-345-1	Sequence 1, Appl
24	44	22.9	329	US-08-884-681-3	Sequence 3, Appl
25	44	22.9	329	US-09-258-643-3	Sequence 3, Appl
26	44	22.9	334	US-09-120-365-64	Sequence 64, Appl
27	44	22.9	334	US-09-515-039-64	Sequence 64, Appl

28	44	22.9	368	4	US-08-818-112-114	Sequence 114, App
29	44	22.9	368	4	US-08-818-111-109	Sequence 109, App
30	44	22.9	368	4	US-09-056-556-114	Sequence 114, App
31	44	22.9	368	4	US-09-072-596-109	Sequence 109, App
32	44	22.9	372	2	US-08-884-681-1	Sequence 1, Appl
33	44	22.9	372	4	US-09-258-643-1	Sequence 1, Appl
34	44	22.9	373	3	US-08-930-966A-11	Sequence 11, Appl
35	44	22.9	476	4	US-09-134-001C-4458	Sequence 4458, Ap
36	44	22.9	761	3	US-09-012-710-13	Sequence 13, Appl
37	44	22.9	761	4	US-09-556-273-13	Sequence 13, Appl
38	43.5	22.7	344	2	US-09-055-097-3	Sequence 3, Appl
39	43.5	22.7	438	1	US-07-973-324A-2	Sequence 2, Appl
40	43.5	22.7	438	1	US-08-343-380-2	Sequence 2, Appl
41	43.5	22.7	438	4	US-09-072-435-2	Sequence 2, Appl
42	43.5	22.7	438	4	US-09-072-917A-2	Sequence 2, Appl
43	43.5	22.7	616	4	US-09-298-367B-11	Sequence 11, Appl
44	43.5	22.7	618	4	US-09-134-001C-3169	Sequence 3169, Ap
45	43.5	22.7	2304	4	US-09-324-867-4	Sequence 4, Appl

#### ALIGNMENTS

```
RESULT 1
US-09-323-873A-20
; Sequence 20, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahen Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; TITLE OF INVENTION: EXPRESSED IN HUMAN CANCERS AND USES THEREOF
; FILE REFERENCE: 129.16USU2
; CURRENT APPLICATION NUMBER: US/09/323,873A
; PRIOR FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-01
; PRIOR APPLICATION NUMBER: 60/091,183
; PRIOR FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: STEAP-1 PEPTIDE
US-09-323-873A-20

Query Match      100.0%; Score 192; DB 4; Length 34;
Best Local Similarity 100.0%; Pred. No. 3.7e-20;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRSRYRYKLIMWAVQVQONKEDAWIEHDVWRMEI 34
DB 1 RRSRYRYKLIMWAVQVQONKEDAWIEHDVWRMEI 34

RESULT 2
US-09-323-873A-2
; Sequence 2, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahen Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Steve Chappell Mitchell
```

TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS  
FILE REFERENCE: 129.16USU2  
CURRENT APPLICATION NUMBER: US/09/323,873A  
CURRENT FILING DATE: 1999-06-01, 520  
PRIOR APPLICATION NUMBER: 60/087,520  
PRIOR FILING DATE: 1998-06-01  
PRIOR APPLICATION NUMBER: 60/091,183  
PRIOR FILING DATE: 1998-06-30  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 339  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-09-323-873A-2

Query Match 100.0%; Score 192; DB 4; Length 339;  
Best Local Similarity 100.0%; Pred. No. 4,9e-19;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRSRYKLLMAYQVQONKEDAMIEHDVWRMEI 34  
DB 185 RRSRYKLLMAYQVQONKEDAMIEHDVWRMEI 218

RESULT 3  
US-09-083-521-1  
Sequence 1, Application US/09083521  
Patent No. 6048970  
GENERAL INFORMATION:  
APPLICANT: Lal, Preeti  
APPLICANT: Guegler, Karl J.  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: PROSTATE GROWTH-ASSOCIATED MEMBRANE PROTEINS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,521  
FILING DATE: Herewith  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: CERONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0527 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 141 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PROSTUT10  
CLONE: 1691243  
US-09-083-521-1

Query Match 55.7%; Score 107; DB 3; Length 141;  
Best Local Similarity 58.8%; Pred. No. 1e-07;  
Matches 20; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 RRSRYKLLMAYQVQONKEDAMIEHDVWRMEI 34  
DB 13 RRSRYKLLMAYQVQONKEDAMIEHDVWRMEI 46

RESULT 4  
US-09-323-873A-8  
Sequence 8, Application US/09323873A  
Patent No. 6329503  
GENERAL INFORMATION:  
APPLICANT: Daniel E. Afar  
APPLICANT: Rene S. Hubert  
APPLICANT: Kahan Leong  
APPLICANT: Arthur B. Raitano  
APPLICANT: Douglas C. Saffran  
APPLICANT: Steve Chappell Mitchell  
TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS  
FILE REFERENCE: 129.16USU2  
CURRENT APPLICATION NUMBER: US/09/323,873A  
CURRENT FILING DATE: 1999-06-01  
PRIOR APPLICATION NUMBER: 60/087,520  
PRIOR FILING DATE: 1998-06-01  
PRIOR APPLICATION NUMBER: 60/091,183  
PRIOR FILING DATE: 1998-06-30  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 8  
LENGTH: 173  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-09-323-873A-8

Query Match 55.7%; Score 107; DB 4; Length 173;  
Best Local Similarity 58.8%; Pred. No. 1.3e-07;  
Matches 20; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 RRSRYKLLMAYQVQONKEDAMIEHDVWRMEI 34  
DB 81 RRSRYKLLMAYQVQONKEDAMIEHDVWRMEI 114

RESULT 5  
US-09-323-873A-32  
Sequence 32, Application US/09323873A  
Patent No. 6329503  
GENERAL INFORMATION:  
APPLICANT: Daniel E. Afar  
APPLICANT: Rene S. Hubert  
APPLICANT: Kahan Leong  
APPLICANT: Arthur B. Raitano  
APPLICANT: Douglas C. Saffran  
APPLICANT: Steve Chappell Mitchell  
TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS  
FILE REFERENCE: 129.16USU2  
CURRENT APPLICATION NUMBER: US/09/323,873A  
CURRENT FILING DATE: 1999-06-01  
PRIOR APPLICATION NUMBER: 60/087,520  
PRIOR FILING DATE: 1998-06-01  
PRIOR APPLICATION NUMBER: 60/091,183  
PRIOR FILING DATE: 1998-06-30  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 32  
LENGTH: 15  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-09-323-873A-32

Query Match 44.3%; Score 85; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 9.1e-06;



```
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 13 YQVQONKEDAWIEH 27
Db 1 YQVQONKEDAWIEH 15

RESULT 6
US-09-149-476-436
; Sequence 436, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23

; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,879
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,880
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,894
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,911
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,636
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,874
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,910
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,864
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,631
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,845
```

EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,892  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,761  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/047,595  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,599  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,588  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,585  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,586  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,590  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,594  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,589  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,593  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,614  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,578  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,576  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/047,501  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,670  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/056,632  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,664  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,876  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,881  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,909  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,875  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,862  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,887  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,908  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/057,650  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/056,884  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,669  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/049,610  
EARLIER FILING DATE: 1997-06-13  
EARLIER APPLICATION NUMBER: 60/061,060  
EARLIER FILING DATE: 1997-10-02

Query Match 26.0%; Score 50; DB 4; Length 370;  
Best Local Similarity 40.7%; Pred. No. 23;  
Matches 11; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 5 RYKLNWYQOVQONKEDAWIEHDVWR 31  
DB 191 ROKUNDALDRLEKEPAFDFDVR 217

RESULT 7  
US-08-560-005-10  
Sequence 10, Application US/08560005  
Patent No. 6001354  
GENERAL INFORMATION:  
APPLICANT: Pot, David A.  
APPLICANT: Williams, Lewis T.  
APPLICANT: Jefferson, Anne Bennett  
TITLE OF INVENTION: No. 6001354e1 Grp2 Associating Protein and Nucleic  
TITLE OF INVENTION: Acids Encoding Therefor  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Townsend and Townsend and Crew  
STREET: One Market Plaza, Steuart Tower, Suite 2000  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/560,005  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Dow, Karen B.  
REGISTRATION NUMBER: 29,684  
REFERENCE/DOCKET NUMBER: 2307K-0624000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 654 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Region  
LOCATION: 1..654  
OTHER INFORMATION: /note= "consensus"  
US-08-560-005-10

Query Match 25.3%; Score 48.5; DB 3; Length 654;  
Best Local Similarity 25.0%; Pred. No. 70;  
Matches 8; Conservative 12; Mismatches 9; Indels 3; Gaps 1;

QY 3 SYRKLNWYQOVQONKEDAWIEHDVWR 31  
DB 348 TYKRGDAVYKOKTSEKTRVPMCDRIWK 379

RESULT 8  
US-09-418-540-10  
Sequence 10, Application US/09418540  
Patent No. 6296848  
GENERAL INFORMATION:  
APPLICANT: Pot, David A.  
APPLICANT: Williams, Lewis T.  
APPLICANT: Jefferson, Anne Bennett  
APPLICANT: Majerus, Philip W.  
TITLE OF INVENTION: No. 6296848e1 Grp2 Associating Protein and Nucleic  
TITLE OF INVENTION: Acids Encoding Therefor  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Townsend and Townsend and Crew  
STREET: One Market Plaza, Steuart Tower, Suite 2000  
CITY: San Francisco  
STATE: California

```
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/418,540
FILING DATE: 14-OCT-1999
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/560,005
FILING DATE: 17-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Dow, Karen B.
REGISTRATION NUMBER: 29,684
REFERENCE/DOCKET NUMBER: 2307K-0624000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 654 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Region
LOCATION: 1..654
OTHER INFORMATION: /note="consensus"
US-09-418-540-10
```

```
Query Match 25.3%; Score 48.5; DB 4; Length 654;
Best Local Similarity 25.0%; Pred. No. 70;
Matches 8; Conservative 12; Mismatches 9; Indels 3; Gaps 1;
```

```
OY 3 SYRYKLNMAVQOQ--VOQNKEDAWIEHDVWR 31
DB 348 TYKYRGDVAAYKQKQSEKTRVPAMCDRLIMK 379
```

```
RESULT 9
US-09-134-001C-3412
Sequence 3412, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3412
LENGTH: 178
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3412
```

```
Query Match 24.7%; Score 47.5; DB 4; Length 178;
Best Local Similarity 41.7%; Pred. No. 22;
Matches 10; Conservative 6; Mismatches 7; Indels 1; Gaps 1;
```

```
OY 1 RRSRYKLNMAVQOQVQONKEDAW 24
DB 139 RQNHKSQSDPMHQONQOHK-DAM 161
```

```
RESULT 10
US-09-428-517-2
Sequence 2, Application US/09428517
Patent No. 6251636
GENERAL INFORMATION:
APPLICANT: Betlach, Mary C.
APPLICANT: Shah, Sanjay Krishnakant
APPLICANT: McDaniel, Robert
APPLICANT: Tang, Li
TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 30062-20029.00
CURRENT APPLICATION NUMBER: US/09/428,517
CURRENT FILING DATE: 1999-10-28
EARLIER APPLICATION NUMBER: 60/120,254
EARLIER FILING DATE: 1999-02-16
EARLIER APPLICATION NUMBER: 60/106,100
EARLIER FILING DATE: 1998-10-29
NUMBER OF SEQ ID NOS: 12
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 2
LENGTH: 4150
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Recombinant
US-09-428-517-2
```

```
Query Match 24.7%; Score 47.5; DB 4; Length 4150;
Best Local Similarity 44.4%; Pred. No. 7,7e+02;
Matches 12; Conservative 4; Mismatches 10; Indels 1; Gaps 1;
```

```
OY 5 RYKLNMAVQOQVQONKEDAWIEHDVWR 31
DB 1924 RYWLDDKAQRRERLED-WRYHVEWR 1949
```

```
RESULT 11
US-09-121-979-4
Sequence 4, Application US/09121979
Patent No. 6159709
GENERAL INFORMATION:
APPLICANT: Korneluk, Robert G.
APPLICANT: Holcik, Martin
APPLICANT: Liston, Peter
TITLE OF INVENTION: XIP IRES AND USES THEREOF
FILE REFERENCE: 07891/021001
CURRENT APPLICATION NUMBER: US/09/121,979
CURRENT FILING DATE: 1998-07-24
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 236
TYPE: PRT
ORGANISM: Homo sapiens
US-09-121-979-4
```

```
Query Match 24.0%; Score 46; DB 4; Length 236;
Best Local Similarity 39.1%; Pred. No. 50;
Matches 9; Conservative 2; Mismatches 6; Indels 6; Gaps 1;
```

```
OY 8 LNNMAVQOQVQONKEDAWIEHDVW 30
DB 46 LANN-----KPKEDPWEQHAKW 62
```

```
RESULT 12
US-09-332-319-4
Sequence 4, Application US/09332319
Patent No. 6171821
GENERAL INFORMATION:
APPLICANT: Korneluk, Robert G.
APPLICANT: Holcik, Martin
```

```

; APPLICANT: Liston, Peter
; TITLE OF INVENTION: XIAP INHIBITORS AND USES THEREOF
; FILE REFERENCE: 07891/021002
; CURRENT APPLICATION NUMBER: US/09/332,319
; CURRENT FILING DATE: 1999-06-14
; EARLIER APPLICATION NUMBER: 09/121,979
; EARLIER FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 4
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-332-319-4

```

```

Query Match      24.0%; Score 46; DB 4; Length 236;
Best Local Similarity 39.1%; Pred. No. 50;
Matches 9; Conservative 2; Mismatches 6; Indels 6; Gaps 1;

```

```

QY      8  LNNWYQVOONKEDAWIEHDVW 30
DB      46  LANN-----KPKEDPWEQHAKW 62

```

```

RESULT 13
; Sequence 2, Application US/09239867.
; Patent No. 6331412
; GENERAL INFORMATION:
; APPLICANT: Robert G. Korneluk et al.
; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR MODULATING
; TITLE OF INVENTION: MALE FERTILITY
; FILE REFERENCE: 07891/018002
; CURRENT APPLICATION NUMBER: US/09/239,867
; CURRENT FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/073,001
; PRIOR FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 2
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-239-867-2

```

```

Query Match      24.0%; Score 46; DB 4; Length 236;
Best Local Similarity 39.1%; Pred. No. 50;
Matches 9; Conservative 2; Mismatches 6; Indels 6; Gaps 1;

```

```

QY      8  LNNWYQVOONKEDAWIEHDVW 30
DB      46  LANN-----KPKEDPWEQHAKW 62

```

```

RESULT 14
; Sequence 2, Application US/08746789A
; Patent No. 5789200
; GENERAL INFORMATION:
; APPLICANT: Ismaili Kola, Martin J. Tyms, Christine DeBouck
; TITLE OF INVENTION: A No. 5789200e1 Human ETS Family Member, ELF3
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road, P.O. Box 1539
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM 486
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS

```

```

; SOFTWARE: MICROSOFT WORD
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,789A
; FILING DATE: No. 5789200e1 15, 1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: William T. Han
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: ATG 50024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610 270 5219
; TELEFAX: 610 270 4026
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 371
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-746-789A-2

```

```

Query Match      23.4%; Score 45; DB 1; Length 371;
Best Local Similarity 47.1%; Pred. No. 1.1e+02;
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

```

```

QY      7  KLNWYQVOONKEDAWIEHDVW 23
DB      71  QVLDWISYQVEKXKXDA 87

```

```

RESULT 15
; Sequence 9, Application US/09342648
; Patent No. 6248584
; GENERAL INFORMATION:
; APPLICANT: Canoon, Rebecca E.
; APPLICANT: Odell, Joan
; APPLICANT: Rafaleki, Antoni
; TITLE OF INVENTION: Transcription Coactivators
; FILE REFERENCE: BB-1169-B
; CURRENT APPLICATION NUMBER: US/09/342,648
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: 60/092,659
; EARLIER FILING DATE: July 13, 1998
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Microsoft Office 97
; SEQ ID NO: 9
; LENGTH: 885
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-342-648-9

```

```

Query Match      23.4%; Score 45; DB 4; Length 885;
Best Local Similarity 37.0%; Pred. No. 3e+02;
Matches 10; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

```

```

QY      2  RSYRYKLNWYQVOONKEDAWIEHD 28
DB      611  RSSYRSLLSAEBAKQKKEKVAHYE 637

```

```

Search completed: March 26, 2003, 16:51:40
Job time : 22.7442 secs

```

GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 16:43:45 ; Search time 35.5814 Seconds  
(without alignments)  
196,889 Million cell updates/sec

Title: US-10-010-667a-20  
Perfect score: 192  
Sequence: 1 RRSRYKLIMWAYQVQONKEDAMIEHDVWMEI 34

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP\_invertebrate:\*  
6: SP\_mammal:\*  
7: SP\_mhc:\*  
8: SP\_organeller:\*  
9: SP\_phage:\*  
10: SP\_plant:\*  
11: SP\_rodent:\*  
12: SP\_virus:\*  
13: SP\_vertebrate:\*  
14: SP\_unclassified:\*  
15: SP\_viruses:\*  
16: SP\_bacteriapi:\*  
17: SP\_archaeapi:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	192	100.0	264	4	Q9Y6U5	Q9Y6U5 homo sapien
2	192	100.0	338	6	Q9GU50	Q9GU50 sus scrofa
3	187	97.4	339	11	Q9CWR7	Q9CWR7 mus musculus
4	187	97.4	339	11	Q924J9	Q924J9 mus musculus
5	139	72.4	339	11	Q924Z2	Q924Z2 mus musculus
6	109	56.8	488	11	Q99P41	Q99P41 rattus norv
7	109	56.8	514	11	Q924Z1	Q924Z1 mus musculus
8	107	55.7	488	4	Q9NVB5	Q9NVB5 homo sapien
9	96.5	50.3	456	4	Q8TDP3	Q8TDP3 homo sapien
10	96.5	50.3	487	4	Q8TF03	Q8TF03 homo sapien
11	62.5	32.6	480	16	O67595	O67595 aquifex aeo
12	60	31.2	926	4	Q8T849	Q8T849 homo sapien
13	60	31.2	926	11	Q8R554	Q8R554 mus musculus
14	58.5	30.5	1027	5	Q95QC7	Q95QC7 caenorhabd
15	57.5	29.9	858	4	Q9N053	Q9N053 homo sapien
16	56	29.2	150	12	Q9YVJ4	Q9YVJ4 melanoplus

17	56	29.2	230	2	Q9RFP3	Q9RFP3 mycoplasma
18	55.5	28.9	502	2	Q48904	Q48904 microcystis
19	53.5	27.9	427	4	Q8WMA7	Q8WMA7 homo sapien
20	53	27.6	152	17	Q30299	Q30299 archaeoglob
21	53	27.6	335	2	Q87519	Q87519 escherichia
22	53	27.6	429	17	Q8TTF0	Q8TTF0 methanobarc
23	52	27.1	132	5	Q91W00	Q91W00 human immun
24	52	27.1	222	5	Q9N7A8	Q9N7A8 leishmania
25	52	27.1	332	5	Q9FHE5	Q9FHE5 arabidopsis
26	52	27.1	1218	5	Q9W296	Q9W296 drosoephila
27	52	27.1	1427	9	Q9WPE5	Q9WPE5 human immun
28	51.5	26.8	169	9	Q9G045	Q9G045 mycobacteri
29	51.5	26.8	657	16	Q8RDS2	Q8RDS2 fusobacteri
30	51.5	26.8	713	16	Q25402	Q25402 helicobacte
31	51.5	26.8	1252	10	Q8W363	Q8W363 oryza sativ
32	51	26.6	90	16	Q8YUH7	Q8YUH7 anabaena sp
33	51	26.6	170	9	Q9R052	Q9R052 mycobacteri
34	51	26.6	289	2	Q8RPU7	Q8RPU7 klebsiella
35	51	26.6	302	15	Q8Q5X2	Q8Q5X2 human immun
36	51	26.6	444	5	Q21449	Q21449 caenorhabd
37	51	26.6	487	15	Q8USU8	Q8USU8 human immun
38	51	26.6	793	16	Q9HZC9	Q9HZC9 pseudomonas
39	50.5	26.3	201	16	Q8YMA3	Q8YMA3 anabaena sp
40	50.5	26.3	220	12	Q91Z87	Q91Z87 indian pean
41	50.5	26.3	402	16	Q8Z040	Q8Z040 anabaena sp
42	50.5	26.3	712	16	Q9ZLE7	Q9ZLE7 helicobacte
43	50	26.0	110	5	Q9W2E8	Q9W2E8 drosoephila
44	50	26.0	204	16	Q8YMT6	Q8YMT6 anabaena sp
45	50	26.0	345	11	Q91YU3	Q91YU3 mus musculu

#### ALIGNMENTS

RESULT 1  
Q9Y6U5 PRELIMINARY; FRT; 264 AA.  
ID Q9Y6U5  
AC Q9Y6U5;  
DT 01-NOV-1999 (T-EMBLrel. 12, Created)  
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
DE WUGSC:H\_RG087B15.1 protein (fragment).  
GN WUGSC:H\_RG087B15.1  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strong C., Layman D., Graves T., Sirowmatt C.;  
RT "The sequence of Homo sapiens BAC clone CTB-87B15."  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Strong C., Layman D., Graves T., Sirowmatt C.;  
RT "Toward a complete human genome sequence."  
RL Genome Res. 8:1097-1108(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Waterston R.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL:AC005061; AAD43182.1; -.  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 264 AA; 31203 MW; 35C9483003557E72 CRC64;  
Query Match 100.0%; Score 192; DB 4; Length 264;  
Best Local Similarity 100.0%; Pred. No. 2,8e-17;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RRSRYKLIMWAYQVQONKEDAMIEHDVWMEI 34  
|||||

DB 195 RRSRYKLLMWAYQVQONKEDAMIEHDVWMEI 228

RESULT 2

Q9GL50 PRELIMINARY; PRT; 338 AA.

AC Q9GL50; (T-EMBLrel. 16, Created)

DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)

DE Six transmembrane endothelial antigen of PAEC.

OS Sus scrofa (pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

OC NCBI\_TaxID=9823;

OX NCBI\_TaxID=9823;

RN [1]

RP SEQUENCE FROM N.A.

RA Nagasaki T., Bouleday G., Coupel S., Coulon F., Tesson L., Healan J.-M., Souillou J.-P., Charreau B.;

RT "Differential gene expression in endothelial cells during TNF-alpha- and LPS-mediated activation."

RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

RL EMBL; AF319659; AAC33868.1; -.

DR Transmembrane.

KW TRANSMEMBRANE.

SQ SEQUENCE 338 AA; 39918 MW; ED490E86E067A32B CRC64;

Query Match 100.0%; Score 192; DB 6; Length 338;

Best Local Similarity 100.0%; Pred. No. 3,7e-17;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRSRYKLLMWAYQVQONKEDAMIEHDVWMEI 34

DB 184 RRSRYKLLMWAYQVQONKEDAMIEHDVWMEI 217

RESULT 3

Q9CWR7 PRELIMINARY; PRT; 339 AA.

AC Q9CWR7; (T-EMBLrel. 17, Created)

DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)

DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)

DE 2410007B19Rik protein.

GN STEAP OR 2410007B19RIK.

OS Mus musculus (mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI\_TaxID=10090;

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I., Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R., Kodora K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H., Kuehl P., Lewis S., Matsuo Y., Nikaide Y., Pesole G., Quackenbush J., Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Guelincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohzuki S., Hayashizaki Y.,

RT "Functional annotation of a full-length mouse cDNA collection."

RL Nature 409:685-690(2001).

DR EMBL; AK010437; BAB26938.1; -.

DR MGD; MGI:1917608; Steap.

SQ SEQUENCE 339 AA; 39264 MW; 3F7AB9C7520F0968 CRC64;

Query Match 97.4%; Score 187; DB 11; Length 339;

Best Local Similarity 94.1%; Pred. No. 1.6e-16;

Matches 32; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRSRYKLLMWAYQVQONKEDAMIEHDVWMEI 34

DB 185 RRSRYKLLMWAYQVQONKEDAMIEHDVWMEI 218

RESULT 4

Q924J9 PRELIMINARY; PRT; 339 AA.

AC Q924J9; (T-EMBLrel. 19, Created)

DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)

DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)

DE Six-transmembrane epithelial antigen of the prostate.

GN STEAP.

OS Mus musculus (mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI\_TaxID=10090;

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=C57BL/6;

RX MEDLINE=21371909; PubMed=11479226;

RA Yang D., Holt G.B., Velders M.P., Kwon E.D., Kaet W.M.;

RT "Murine six-transmembrane epithelial antigen of the prostate, prostate stem cell antigen, and prostate-specific membrane antigen: prostate-specific cell-surface antigens highly expressed in prostate cancer of transgenic adenocarcinoma mouse prostate mice."

RT Cancer Res. 61:5857-5860(2001).

RL EMBL; AF297098; AA83126.1; -.

DR MGD; MGI:1917608; Steap.

KW TRANSMEMBRANE.

SQ SEQUENCE 339 AA; 39318 MW; 4B26A71FF559E84F CRC64;

Query Match 97.4%; Score 187; DB 11; Length 339;

Best Local Similarity 94.1%; Pred. No. 1.6e-16;

Matches 32; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRSRYKLLMWAYQVQONKEDAMIEHDVWMEI 34

DB 185 RRSRYKLLMWAYQVQONKEDAMIEHDVWMEI 218

RESULT 5

Q924Z2 PRELIMINARY; PRT; 339 AA.

AC Q924Z2; (T-EMBLrel. 19, Created)

DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)

DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)

DE Ductal.

GN 1010001D01RIK.

OS Mus musculus (mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI\_TaxID=10090;

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Serru V., Manivet P., Lambin D., Vaubourville M., Kellermann O., Loric S.;

RT "Prostate and non-prostate expression of ductal, the mouse ortholog of human STEAP."

RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

RL EMBL; AY029584; AAK50537.1; -.

DR MGD; MGI:1915678; 1010001D01RIK.

DR MGD; MGI:1915678; 1010001D01RIK.

SQ SEQUENCE 339 AA; 39109 MW; 32A2C29F2E33B5D0 CRC64;

Query Match 72.4%; Score 139; DB 11; Length 339;  
 Best Local Similarity 73.5%; Pred. No. 3e-10;  
 Matches 25; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 RRSRYKLLNMAVOOQONKEDAMIEHDVVRMEI 34  
 DB 185 RRSRYDLVNLAVKQVLANKSRLWVEEVRMEI 218

## RESULT 6

Q99P41 PRELIMINARY; PRT; 488 AA.

AC Q99P41; 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Tumor suppressor pHyde.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DUNNING; TISSUE=PROSTATE CANCER TUMOR;  
 RX MEDLINE=20424188; PubMed=10969787;  
 RA Steiner M.S., Zhang X., Wang Y., Lu Y.;  
 RT "Growth inhibition of prostate cancer by an adenovirus expressing a  
 novel tumor suppressor gene, pHyde.";  
 RL Cancer Res. 60:4419-4425(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DUNNING; TISSUE=PROSTATE CANCER TUMOR;  
 RA Lu Y., Rinaldy A.R., Steiner M.S.;  
 RN Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Steiner M.S., Rinaldy A.;  
 RT "A novel prostate cancer associated gene.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF335281; AAK00361.1; -;  
 DR EMBL; AF238865; AAL78207.1; -;  
 SQ SEQUENCE 488 AA; 54640 MW; 59FF0721919PDCB CRC64;

Query Match 56.8%; Score 109; DB 11; Length 488;  
 Best Local Similarity 58.8%; Pred. No. 3.6e-06;  
 Matches 20; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 RRSRYKLLNMAVOOQONKEDAMIEHDVVRMEI 34  
 DB 326 RRSRYDLVNLAVKQVLANKSRLWVEEVRMEI 359

## RESULT 7

Q92421 PRELIMINARY; PRT; 514 AA.

AC Q92421; 01-JUN-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Dudulin 2.  
 GN 1010001D01RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Serru V., Lambin D., Lenoir C., Manivet P., Vaubourdel M.,  
 RA Kellermann O., Loric S.;  
 RT "Molecular cloning and expression of mouse dudulin 2.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY029586; AAK50539.1; -;  
 DR MGD; MGI:1915678; 1010001D01RIK.

SQ SEQUENCE 514 AA; 57268 MW; 33986C288AEC0E2 CRC64;  
 Query Match 56.8%; Score 109; DB 11; Length 514;  
 Best Local Similarity 58.8%; Pred. No. 3.8e-06;  
 Matches 20; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 RRSRYKLLNMAVOOQONKEDAMIEHDVVRMEI 34  
 DB 326 RRSRYDLVNLAVKQVLANKSRLWVEEVRMEI 359

## RESULT 8

Q9NVB5 PRELIMINARY; PRT; 488 AA.

AC Q9NVB5; 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE CDNA Fd10829 f18, clone NT2RP4001138 (Dudulin 2).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,  
 RA Matsunawa H., Ishii S., Kawai Y., Satou K., Yamamoto J., Wakamatsu A.,  
 RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;  
 RT "NEO human cDNA sequencing project.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Serru V., Manivet P., Lenoir C., Bachwege P., Lambin D.,  
 RA Vaubourdel M., Kellermann O., Loric S.;  
 RT "Dudulin 2, a new tumor antigen expressed in various human tumors.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK001691; BAA91839.1; -;  
 DR EMBL; AY029585; AAK50538.1; -;  
 DR InterPro; IPR003006; IG\_MHC.  
 DR PROSITE; PS00290; IG\_MHC; UNKNOWN.1.  
 SQ SEQUENCE 488 AA; 54616 MW; BC0BCA483335AAD6 CRC64;

Query Match 55.7%; Score 107; DB 4; Length 488;  
 Best Local Similarity 55.9%; Pred. No. 6.5e-06;  
 Matches 19; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 RRSRYKLLNMAVOOQONKEDAMIEHDVVRMEI 34  
 DB 326 RRAHRYDLVNLAVKQVLANKSHLWVEEVRMEI 359

## RESULT 9

Q8TDP3 PRELIMINARY; PRT; 456 AA.

AC Q8TDP3; 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE PHYDE II.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wang C., Alay J.A., Steiner M.S.;  
 RT "Second human member of pHyde family, Human pHyde II.";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF262322; AAM08128.1; -;  
 SQ SEQUENCE 456 AA; 50430 MW; C5F7C7008D5521E CRC64;

Query Match 50.3%; Score 96.5; DB 4; Length 456;

Best Local Similarity 55.9%; Pred. No. 0.00014;  
Matches 19; Conservative 6; Mismatches 8; Indels 1; Gaps 1;

QY 1 RRSRYKLMAYQVOQNKEDAMIEHDVWRMEI 34  
DB 326 RRAHRYDLVNLAVKQVLANKSHLWVE-EVWRMEI 358

## RESULT 10

Q8TF03 PRELIMINARY; PRT; 487 AA.

AC Q8TF03; 01-JUN-2002 (TReMBLrel. 21, Created)  
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Tumor suppressor pHyde.

OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxId=9606;

RP SEQUENCE FROM N.A.  
RA Steiner M.S., Alley J.A., Wang C.;

RT "A novel prostate-derived tumor suppressor";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF238864; AAL78206.1; -

SQ SEQUENCE 487 AA; 54402 MW; AFI60535906F68 CRC64;

Query Match 50.3%; Score 96.5; DB 4; Length 487;  
Best Local Similarity 55.9%; Pred. No. 0.00015;  
Matches 19; Conservative 6; Mismatches 8; Indels 1; Gaps 1;

QY 1 RRSRYKLMAYQVOQNKEDAMIEHDVWRMEI 34  
DB 326 RRAHRYDLVNLAVKQVLANKSHLWVE-EVWRMEI 358

## RESULT 11

Q67595 PRELIMINARY; PRT; 480 AA.

AC Q67595; 01-AUG-1998 (TReMBLrel. 07, Created)  
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
DE Hypothetical protein A0\_1687.

GN A0\_1687.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
OX NCBI\_TaxId=63363;

RP SEQUENCE FROM N.A.  
RC STRAIN=V5;

RX MEDLINE=98196666; Pubmed=9537320;  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
RA Graham D.E., Overbeek R., Sneed G.J., Keller M., Aubay M., Huber R.,  
RA Feldman R.A., Short J.M., Olson M.G., Swanson R.V.;  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
RT aeolicus";  
RL Nature 382:353-358(1998).

KW EMBL; AB000751; AAC07560.1; -  
DR Hypothetical protein; Complete proteome.  
SQ SEQUENCE 480 AA; 55582 MW; 7BED49517A218046 CRC64;

Query Match 32.6%; Score 62.5; DB 16; Length 480;  
Best Local Similarity 25.0%; Pred. No. 4;  
Matches 12; Conservative 17; Mismatches 4; Indels 15; Gaps 2;

QY 1 RRSRYKLMAYQVOQNKEDAMIEHDVWRMEI 33  
DB 10 RRTYRLIPAKVSDIGKEKYLWDSYEGFRLEKSEDFEKKYKX 57

## RESULT 12

Q8TE49 PRELIMINARY; PRT; 926 AA.

AC Q8TE49; 01-JUN-2002 (TReMBLrel. 21, Created)  
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Cezanne 2 protein.

GN CEZANNE 2.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxId=9606;

RP SEQUENCE FROM N.A.  
RA Evans P.C., Coadwell W.J., Kilshaw P.J.;

RT "Isolation of a novel human gene, Cezanne 2";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ430383; CAD23047.1; -

SQ SEQUENCE 926 AA; 100676 MW; 6E4623C2EB2C8058 CRC64;

Query Match 31.2%; Score 60; DB 4; Length 926;  
Best Local Similarity 64.3%; Pred. No. 17;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 11 WAYQVOQNKEDAW 24  
DB 250 WRWQQTQONKEEW 263

## RESULT 13

Q8R554 PRELIMINARY; PRT; 926 AA.

AC Q8R554; 01-JUN-2002 (TReMBLrel. 21, Created)  
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Cezanne 2 protein.

GN CEZANNE 2.  
OS Mus musculus (mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;

RP SEQUENCE FROM N.A.  
RC STRAIN=C57 BL/6;  
RA Evans P.C., Coadwell W.J., Kilshaw P.J.;

RT "Isolation of a novel murine gene, Cezanne 2";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ430384; CAD23048.1; -

SQ SEQUENCE 926 AA; 100796 MW; 4D6BD05A0410BED9 CRC64;

Query Match 31.2%; Score 60; DB 11; Length 926;  
Best Local Similarity 64.3%; Pred. No. 17;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 11 WAYQVOQNKEDAW 24  
DB 252 WRWQQTQONKEEW 265

## RESULT 14

Q950C7 PRELIMINARY; PRT; 1027 AA.

AC Q950C7; 01-DEC-2001 (TReMBLrel. 19, Created)  
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
DE Putative TRP homologous cation channel protein (Fragment).  
GN LTRPC4/CED-11.

OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxId=6239;







GenCore version 5.1.4\_p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 16:38:14 ; Search time 66.0233 Seconds  
(without alignments)  
68.620 Million cell updates/sec

Title: US-10-010-667a-20

Perfect score: 192  
Sequence: 1 RRSRYKLLMAYQQVQNKEDAWIHDVWMEI 34

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_101002:\*

- 1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	192	100.0	34	21	AA158199 Human STRAP-1 pept
2	192	100.0	34	22	AA502787 Extracellular loop
3	192	100.0	129	22	AA875314 Gene 20 human secr
4	192	100.0	129	22	AA875315 Human secreted pro
5	192	100.0	339	21	AA158194 Human STRAP-1 prot
6	192	100.0	339	22	AA158197 Human prostate cDN
7	192	100.0	339	22	AA178845 Human protein SEQ
8	192	100.0	339	22	AA178845 Human protein SEQ
9	192	100.0	339	23	ABG61813 P789P amino acid s
10	192	100.0	339	23	ABG61813 P789P amino acid s
					Human P789P protei

11	192	100.0	375	22	AAE02780
12	182	94.8	339	20	AAW86309
13	179	93.2	374	22	AAW79829
14	109	56.8	488	22	AA849483
15	109	56.8	526	23	AB883365
16	107	55.7	141	21	AA152589
17	107	55.7	173	21	AA158195
18	107	55.7	237	22	AAU04565
19	107	55.7	419	22	AAU10189
20	107	55.7	450	22	AAE02841
21	107	55.7	454	22	AAU10188
22	107	55.7	454	22	AAE02781
23	107	55.7	454	23	ABG61933
24	107	55.7	454	23	AAU80190
25	107	55.7	454	23	AAU76538
26	107	55.7	488	22	AAU10220
27	107	55.7	488	22	AAU04564
28	107	55.7	488	22	AA885775
29	107	55.7	488	22	AA893224
30	107	55.7	488	22	AB883366
31	107	55.7	490	22	AAU10187
32	106	55.2	33	22	AAE02790
33	96.5	50.3	456	22	AA849482
34	96.5	50.3	487	22	AA849481
35	85	44.3	15	21	AA158202
36	85	44.3	15	21	AAE02785
37	57	29.7	128	21	AA158197
38	52.5	27.3	697	22	ABG25300
39	52	27.1	38	22	AB838169
40	52	27.1	38	22	AB823372
41	52	27.1	38	22	AA158802
42	52	27.1	38	22	AAW71318
43	52	27.1	38	22	AAW18991
44	52	27.1	38	22	AAW31599
45	52	27.1	38	23	ABG41121

#### ALIGNMENTS

RESULT 1	
AA158199	AA158199 standard; peptide; 34 AA.
XX	
AC	AA158199;
XX	
DT	14-MAR-2000 (first entry)
XX	
DE	Human STRAP-1 peptide, corresponding to STRAP-1 extracellular region 2.
XX	
KW	Serpentine transmembrane antigen of the prostate; STRAP-1; prostate;
KW	transmembrane domain; type IIIa membrane protein; expression; cancer;
KW	prostate cancer; bladder cancer; colon cancer; pancreatic cancer;
KW	ovarian cancer; tumour antigen; immunisation; immune response;
KW	cellular; humoral; anticancer vaccine; antibody; detection; diagnosis;
KW	prognosis; monitoring; susceptibility; therapeutic inhibitor;
KW	drug targeting; recombinant protein.
XX	
OS	Synthetic.
XX	
XX	Homo sapiens.
PN	W09962941-A2.
XX	
PD	09-DEC-1999.
XX	
PF	01-JUN-1999; 99MO-US12157.
XX	
PR	01-JUN-1998; 98US-0087520.
PR	30-JUN-1998; 98US-0091183.
XX	
PA	(UROC-) UROGENESYS INC.
PA	(AFAR/) AFAR D E.
PA	(HUBE/) HUBERT R. S.

Human six transmem  
Kidney injury asso  
Human protein SEQ  
Rat p-HYDE. Rattu  
Murine Tumour Supp  
Human prostate gro  
Human STRAP-2 prot  
Human G-protein co  
Human ORF3 of Six-  
Human STRAP-2 prot  
Human ORF2 of Six-  
Human six transmem  
Prostate cancer-as  
Human PUMCn prote  
Tumour-associated  
Human Six-Transmem  
Human G-protein co  
Human drug metabol  
Human Tumour Supp  
Human Six-Transmem  
Extracellular loop  
Human p-HYDE 40.  
Human p-HYDE. Hom  
Human STRAP-1 pept  
Peptide 2 of human  
Human STRAP-4 prot  
Novel human diapo  
Peptide #5675 enco  
Protein #5371 enco  
Human brain expres  
Human bone marrow  
Peptide #5425 enco  
Peptide #5636 enco  
Human peptide enco

PA (LEON/) LEONG K.  
 PA (RAIT/) RAITANO A B.  
 PA (SAFR/) SAFFRAN D C.  
 PI Afar DE, Hubert RS, Leong K, Raitano AB, Saffran DC;  
 XX WPI; 2000-072832/06.  
 DR  
 XX Novel proteins useful as diagnostic markers and therapeutic targets,  
 PT particularly for prostatic cancer -  
 XX  
 PS Disclosure, Page 22; 83pp; English.  
 XX  
 CC Sequences AAY58198-Y58200 represent synthetic peptides that correspond  
 CC to the extracellular regions of STRAP-1 (serpentine transmembrane  
 CC antigen of the prostate, AAY58194). These peptides were used to raise  
 CC monoclonal anti-STRAP-1 antibodies. STRAP-1 is the prototype  
 CC member of the STRAP family of proteins (AAY58194-Y58197) which  
 CC exhibit a high degree of structural conservation, but which show  
 CC no significant structural homology to known human proteins. The STRAP-1  
 CC gene has been localised to chromosome 7p22. STRAP-1 is thought to be a  
 CC type IIa membrane protein and is expressed predominantly in prostate  
 CC cells in normal human tissues. Structurally, STRAP-1 is a 339 amino  
 CC acid protein characterised by six transmembrane domains and  
 CC intracellular N- and C-termini, suggesting that it folds in a  
 CC "serpentine" manner into three extracellular and two intracellular loops.  
 CC STRAP-1 mRNA and protein expression is maintained at high levels and  
 CC throughout all stages of prostate cancer. STRAP-1 mRNA and/or protein is  
 CC also overexpressed in certain other cancers, including bladder, colon,  
 CC pancreatic and ovarian cancer. The function of the STRAP proteins is not  
 CC known. They may be ion channels (from the presence of six transmembrane  
 CC domains, a feature which is shared by certain ion channels) or  
 CC GAP-junction proteins (from immunohistochemical staining). STRAP-1 and  
 CC STRAP-2 are cell-surface tumour antigens. Immunisation with a STRAP  
 CC protein induces cellular and humoral immune responses against  
 CC STRAP-expressing cells. STRAP proteins may be used to identify  
 CC specific-binding agents, to produce anticancer vaccines and to generate  
 CC specific antibodies. The antibodies may be used for detection, prognosis,  
 CC and monitoring of cancers (or susceptibility to cancer), as therapeutic  
 CC inhibitors or to target therapeutic agents to their site of action. STRAP  
 CC nucleic acids may be used for recombinant protein production, as  
 CC diagnostic and prognostic reagents, for identifying STRAP-expressing  
 CC cells for screening inhibitors of STRAP expression and for therapeutic  
 CC modulation/inhibition of STRAP expression. Since high levels of STRAP  
 CC proteins are exposed on the cell surface, they are easily targeted by  
 CC systemically administered agents, and because they are expressed mainly  
 CC on prostatic epithelial cells, agents targeted to them should have  
 CC minimal side effects on other tissues.  
 CC  
 XX  
 SQ Sequence 34 AA;  
 Query Match 100.0%; Score 192; DB 21; Length 34;  
 Best Local Similarity 100.0%; Pred. No. 3, 7e-19;  
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RRSRYKTLNMAVYQVQNKEDAWIEHDVVRMEI 34  
 DB 1 RRSRYKTLNMAVYQVQNKEDAWIEHDVVRMEI 34  
 RESULT 2  
 AAE02787  
 ID AAE02787 standard; peptide; 34 AA.  
 XX  
 AC AAE02787;  
 DT 06-AUG-2001 (first entry)  
 XX  
 DE Extracellular Loop #2 of human STRAP-1, suitable for cloning into pFC.  
 XX  
 KM Human; cytostatic; antiproliferative; vaccine; gene therapy;  
 KM six transmembrane epithelial antigen of the prostate-1; STRAP-1;  
 KM chromosome 7p22.3; cancer; prostate; colon; bladder; pancreatic;

KW ovarian; lung; extracellular loop; serpentine transmembrane antigen.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200140276-A2.  
 XX  
 PD 07-JUN-2001.  
 XX  
 PF 06-DEC-2000; 2000WO-US33040.  
 XX  
 PR 06-DEC-1999; 99US-0455486.  
 XX  
 PA (UROG-) UROGENESYS INC.  
 XX  
 PI Afar DEH, Hubert RS, Raitano AB, Saffran DC, Mitchell SC, Faris M;  
 PI Jakobovits A;  
 XX  
 DR WPI; 2001-367804/38.  
 XX  
 PT New STRAP (six transmembrane epithelial antigen of the prostate)  
 PT proteins, expressed in human cancers, useful for detecting and treating  
 PT cancer -  
 XX  
 PS Example 19; Page 102; 187pp; English.  
 XX  
 CC The present invention relates to human six transmembrane epithelial  
 CC antigen of the prostate (STRAP) protein. STRAP is a member of cell  
 CC surface serpentine transmembrane antigens. STRAP gene is used in gene  
 CC therapy. Inhibiting the development or progression of a cancer (eg.  
 CC prostate, colon, bladder, lung, ovarian and pancreatic) expressing STRAP  
 CC or inhibiting growth or killing cells expressing STRAP in a patient,  
 CC comprises administering a vaccine composition to the patient. Treating  
 CC a patient with a cancer that expresses STRAP, or inhibiting growth or  
 CC killing cells expressing STRAP, comprises administering to the patient a  
 CC vector encoding single chain monoclonal antibody that comprises the  
 CC variable domains of the heavy and light chains of the monoclonal antibody  
 CC that specifically binds to STRAP, such that the vector delivers the  
 CC single chain monoclonal antibody coding sequence to the cancer cells and  
 CC the encoded single chain monoclonal antibody is expressed  
 CC intracellularly. The present sequence is extracellular loop of STRAP-1  
 CC suitable for cloning into pFC, which is used in the invention. STRAP-1  
 CC gene is located on chromosome 7p22.3.  
 CC  
 XX  
 SQ Sequence 34 AA;  
 Query Match 100.0%; Score 192; DB 22; Length 34;  
 Best Local Similarity 100.0%; Pred. No. 3, 7e-19;  
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RRSRYKTLNMAVYQVQNKEDAWIEHDVVRMEI 34  
 DB 1 RRSRYKTLNMAVYQVQNKEDAWIEHDVVRMEI 34  
 RESULT 3  
 AAB75314  
 ID AAB75314 standard; Protein; 129 AA.  
 XX  
 AC AAB75314;  
 XX  
 DE 03-APR-2001 (first entry)  
 XX  
 DE Gene 20 human secreted protein homologous amino acid sequence #133.  
 XX  
 KM Human; immunosuppressive; antiarthritic; anti-rheumatic; nootropic;  
 KM antiproliferative; cytostatic; cardiatic; vasotropic; cerebroprotective;  
 KM neuroprotective; antibacterial; virucide; fungicide; ophtalmatological;  
 KM vulnerary; autoimmune disease; hyperproliferative disorder; cancer;  
 KM cardiovascular disorder; cerebrovascular disorder; infection;  
 KM nervous system disorder; ocular disorder; chemotaxis; food additive;  
 KM secreted protein.  
 XX  
 OS Homo sapiens.

XX WO200077021-A1.  
 XX 21-DEC-2000.  
 XX 01-JUN-2000; 2000MO-US15135.  
 XX 11-JUN-1999; 99US-0138632.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX (ROSE/) ROSEN C A.  
 XX Rosen CA, Ruben SM, Komatsoulis GA;  
 XX MPI; 2001-071257/08.  
 XX Nucleic acid molecules encoding human secreted proteins, used in  
 XX preventing, treating or ameliorating a disorder, e.g. Alzheimer's and  
 XX Parkinson's diseases and cancers -  
 XX Disclosure; Page 53-54; 530pp; Englishh.

CC This invention relates to polynucleotide sequences AAF63789 - AAF63836  
 CC which encode human secreted proteins AAB75260 - AAB75287. Included in the  
 CC invention are protein sequences AAB75288 - AAB75341 which are fragments  
 CC of the secreted proteins and amino acid sequences with which these  
 CC fragments share homology. Examples of the activities of the proteins and  
 CC polynucleotides and the activities of their agonists and antagonists  
 CC include; immunosuppressive; antiarthritic; antineumatic;  
 CC antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
 CC neurotropic; neuroprotective; antibacterial; virucide; fungicide;  
 CC ophthalmological; and vulnerary activity. The protein and polynucleotide  
 CC sequences, their agonists and antagonists may be useful for treating,  
 CC preventing and diagnosing diseases and disorders such as autoimmune  
 CC diseases e.g. Rheumatoid arthritis, hyperproliferative disorders  
 CC e.g. neoplasms of the breast or liver, cardiovascular disorders  
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,  
 CC angioneuromatosis, nervous system disorders e.g. Alzheimer's disease,  
 CC infections caused by bacteria, viruses and fungi and ocular disorders  
 CC e.g. corneal infection. The polypeptides can also be used to aid wound  
 CC healing and epithelial cell proliferation, to prevent skin aging due to  
 CC sunburn, to maintain organs before transplantation, for supporting cell  
 CC culture of primary tissues, to regenerate tissues and in chemotaxis. The  
 CC polypeptides can also be used as a food additive or preservative to  
 CC increase or decrease storage capabilities. Included in the invention are  
 CC oligonucleotides AAF63780 - AAF63788 and peptide AAB75239 which are used  
 CC in the identification and characterisation of the DNA and protein  
 CC sequences of the invention.

CC Sequence 129 AA;  
 CC  
 CC Query Match 100.0%; Score 192; DB 22; Length 129;  
 CC Best Local Similarity 100.0%; Pred. No. 1.7e-18;  
 CC Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRSRYRLNMAVQVOONKEDAWIEHDVWRMEI 34  
 Db 60 RRSRYRLNMAVQVOONKEDAWIEHDVWRMEI 93

RESULT 4  
 AAB75315  
 ID AAB75315 standard; Protein; 129 AA.  
 XX AAB75315;  
 XX  
 XX 03-APR-2001 (first entry)  
 XX  
 XX Human secreted protein sequence encoded by gene 20 SEQ ID NO:134.  
 XX  
 XX Human; immunosuppressive; antiarthritic; antineumatic; neurotropic;  
 XX antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
 XX neuroprotective; antibacterial; virucide; fungicide; ophthalmological;

KW vulnerary; autoimmune disease; hyperproliferative disorder; cancer;  
 KW cardiovascular disorder; cerebrovascular disorder; infection;  
 KW nervous system disorder; ocular disorder; chemotaxis; food additive;  
 KW secreted protein.  
 XX Homo sapiens.  
 XX WO200077021-A1.  
 XX 21-DEC-2000.  
 XX 01-JUN-2000; 2000MO-US15135.  
 XX 11-JUN-1999; 99US-0138632.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX (ROSE/) ROSEN C A.  
 XX Rosen CA, Ruben SM, Komatsoulis GA;  
 XX MPI; 2001-071257/08.  
 XX N-PSDB; AAF63808.  
 XX Nucleic acid molecules encoding human secreted proteins, used in  
 XX preventing, treating or ameliorating a disorder, e.g. Alzheimer's and  
 XX Parkinson's diseases and cancers -  
 XX Disclosure; Page 53-54; 530pp; Englishh.

CC This invention relates to polynucleotide sequences AAF63789 - AAF63836  
 CC which encode human secreted proteins AAB75260 - AAB75287. Included in the  
 CC invention are protein sequences AAB75288 - AAB75341 which are fragments  
 CC of the secreted proteins and amino acid sequences with which these  
 CC fragments share homology. Examples of the activities of the proteins and  
 CC polynucleotides and the activities of their agonists and antagonists  
 CC include; immunosuppressive; antiarthritic; antineumatic;  
 CC antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
 CC neurotropic; neuroprotective; antibacterial; virucide; fungicide;  
 CC ophthalmological; and vulnerary activity. The protein and polynucleotide  
 CC sequences, their agonists and antagonists may be useful for treating,  
 CC preventing and diagnosing diseases and disorders such as autoimmune  
 CC diseases e.g. Rheumatoid arthritis, hyperproliferative disorders  
 CC e.g. neoplasms of the breast or liver, cardiovascular disorders  
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,  
 CC angioneuromatosis, nervous system disorders e.g. Alzheimer's disease,  
 CC infections caused by bacteria, viruses and fungi and ocular disorders  
 CC e.g. corneal infection. The polypeptides can also be used to aid wound  
 CC healing and epithelial cell proliferation, to prevent skin aging due to  
 CC sunburn, to maintain organs before transplantation, for supporting cell  
 CC culture of primary tissues, to regenerate tissues and in chemotaxis. The  
 CC polypeptides can also be used as a food additive or preservative to  
 CC increase or decrease storage capabilities. Included in the invention are  
 CC oligonucleotides AAF63780 - AAF63788 and peptide AAB75239 which are used  
 CC in the identification and characterisation of the DNA and protein  
 CC sequences of the invention.

CC Sequence 129 AA;  
 CC  
 CC Query Match 100.0%; Score 192; DB 22; Length 129;  
 CC Best Local Similarity 100.0%; Pred. No. 1.7e-18;  
 CC Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRSRYRLNMAVQVOONKEDAWIEHDVWRMEI 34  
 Db 60 RRSRYRLNMAVQVOONKEDAWIEHDVWRMEI 93

RESULT 5  
 AA58194  
 ID AA58194 standard; Protein; 339 AA.  
 XX AA58194;  
 XX

DT 14-MAR-2000 (first entry)  
 XX Human STRAP-1 protein.  
 XX Serpentine transmembrane antigen of the prostate; STRAP-1; prostate;  
 KW transmembrane domain; type IIIa membrane protein; expression; cancer;  
 KW prostate cancer; bladder cancer; colon cancer; pancreatic cancer;  
 KW ovarian cancer; tumour antigen; immunisation; immune response;  
 KW cellular; humoral; anticancer vaccine; antibody; detection; diagnosis;  
 KW prognosis; monitoring; susceptibility; therapeutic inhibitor;  
 KW drug targeting; recombinant protein.  
 XX Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..69 /note= "Intracellular region 1"  
 FT Domain 70..91 /note= "Intracellular region 1"  
 FT /note= "Transmembrane domain 1"  
 FT Region 92..113 /note= "Extracellular region 1 (AAV58198)"  
 FT Domain 114..136 /note= "Transmembrane domain 2"  
 FT Region 137..162 /note= "Intracellular region 2"  
 FT Domain 163..184 /note= "Transmembrane domain 3"  
 FT Region 185..218 /note= "Extracellular region 2 (AAV58199)"  
 FT Domain 219..241 /note= "Transmembrane domain 4"  
 FT Region 242..253 /note= "Intracellular region 3"  
 FT Domain 252..276 /note= "Transmembrane domain 5"  
 FT Region 277..291 /note= "Extracellular region 3 (AAV58200)"  
 FT Domain 292..313 /note= "Transmembrane domain 6"  
 FT Region 314..339 /note= "Intracellular region 4"  
 FT /note= "Intracellular region 4"  
 XX  
 PN WO962941-A2.  
 XX  
 PD 09-DEC-1999.  
 XX  
 PF 01-JUN-1999; 99WO-US12157.  
 XX  
 PR 01-JUN-1998; 98US-0087520.  
 PR 30-JUN-1998; 98US-0091183.  
 XX  
 PA (UROC-) UROGENESYS INC.  
 PA (AFAR/) AFAR D E.  
 PA (HUBE/) HUBERT R S.  
 PA (LEON/) LEONG K.  
 PA (RAIT/) RAITANO A B.  
 PA (SAFF/) SAFFRAN D C.  
 XX  
 PI Afar DE, Hubert RS, Leong K, Raitano AB, Saffran DC;  
 XX  
 DR WPI; 2000-072832/06.  
 DR N-PsDB; AAZ49395, AAZ49396.  
 XX  
 PT Novel proteins useful as diagnostic markers and therapeutic targets,  
 PT particularly for prostatic cancer -  
 XX  
 PS Claim 1; Fig 1A; 83pp; English.  
 XX  
 CC This sequence represents a novel human protein, STRAP-1 (serpentine  
 CC transmembrane antigen of the prostate). STRAP-1 is the prototype  
 CC member of the STRAP family of proteins (AAV58194-Y58197) which  
 CC exhibit a high degree of structural conservation, but which show  
 CC no significant structural homology to known human proteins. The STRAP-1

CC gene has been localised to chromosome 7p22. STRAP-1 is thought to be a  
 CC type IIIa membrane protein and is expressed predominantly in prostate  
 CC cells in normal human tissues. Structurally, STRAP-1 is a 339 amino  
 CC acid protein characterised by six transmembrane domains and  
 CC intracellular N- and C-termini, suggesting that it folds in a  
 CC "serpentine" manner into three extracellular and two intracellular loops.  
 CC STRAP-1 mRNA and protein expression is maintained at high levels and  
 CC throughout all stages of prostate cancer. STRAP-1 mRNA and/or protein is  
 CC also overexpressed in certain other cancers, including bladder, colon,  
 CC pancreatic and ovarian cancer. The function of the STRAP proteins is not  
 CC known. They may be ion channels (from the presence of six transmembrane  
 CC domains, a feature which is shared by certain ion channels) or  
 CC gap-junction proteins (from immunohistochemical staining). STRAP-1 and  
 CC STRAP-2 are cell-surface tumour antigens. Immunisation with a STRAP  
 CC protein induces cellular and humoral immune responses against  
 CC STRAP-expressing cells. STRAP proteins may be used to identify  
 CC specific-binding agents, to produce anticancer vaccines and to generate  
 CC specific antibodies. The antibodies may be used for detection, prognosis,  
 CC and monitoring of cancers (or susceptibility to cancer), as therapeutic  
 CC inhibitors or to target therapeutic agents to their site of action. STRAP  
 CC nucleic acids may be used for recombinant protein production, as  
 CC diagnostic and prognostic reagents, for identifying STRAP-expressing  
 CC cells for screening inhibitors of STRAP expression and for therapeutic  
 CC modulation/inhibition of STRAP expression. Since high levels of STRAP  
 CC proteins are exposed on the cell surface, they are easily targeted by  
 CC systemically administered agents, and because they are expressed mainly  
 CC on prostatic epithelial cells, agents targeted to them should have  
 CC minimal side effects on other tissues.

XX Sequence 339 AA;

Query Match 100.0%; Score 192; DB 21; Length 339;  
 Best Local Similarity 100.0%; Pred. No. 5,1e-18;  
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RRSRYKLLNMAVQVQONKEDAMIEHDVWRMEI 34  
 Db 185 RRSRYKLLNMAVQVQONKEDAMIEHDVWRMEI 218

RESULT 6  
 AAU69927  
 ID AAU69927 standard; Protein; 339 AA.  
 XX  
 AC AAU69927;  
 XX

DT 30-JUN-2002 (first entry)

XX Human prostate cDNA encoded protein #72.

XX Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.

XX Homo sapiens.  
 XX  
 PN WO200173032-A2.  
 XX  
 PD 04-OCT-2001.  
 XX

PF 27-MAR-2001; 2001MO-US09919.  
 XX

XX 27-MAR-2000; 2000US-0536857.  
 PR 09-MAY-2000; 2000US-0568100.  
 PR 12-MAY-2000; 2000US-0570737.  
 PR 13-JUN-2000; 2000US-0593793.  
 PR 27-JUN-2000; 2000US-0605783.  
 PR 10-AUG-2000; 2000US-0636215.  
 PR 29-AUG-2000; 2000US-0651236.  
 PR 06-SEP-2000; 2000US-0657279.  
 PR 02-OCT-2000; 2000US-0679426.  
 PR 10-OCT-2000; 2000US-0685166.  
 XX

(CORI-) CORIXA CORP.

XX

PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;  
 PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;  
 PI Li SX, Wang A, Skelky YAW, Hepler WT, Henderson RA;  
 XX  
 DR WPI: 2001-639322/73.  
 DR N-PsDB; AAS64160.  
 XX  
 PT New human prostate-specific polypeptides and polynucleotides useful for  
 PT the diagnosis and treatment of cancer, especially prostate cancer -  
 XX  
 PS Claim 2; Page 549; 579pp; English.  
 XX  
 CC The invention relates to isolated prostate-specific  
 CC polynucleotides, polypeptides, fusion proteins of the polypeptides,  
 CC antibodies raised against the polypeptides (or antigenic epitopes  
 CC derived from them) and antigen-presenting cells expressing the  
 CC polypeptides. The antibodies are useful for detecting the presence of  
 CC cancer, especially prostate cancer. The polypeptides, polynucleotides and  
 CC the antigen-presenting cells are useful for stimulating and/or expanding  
 CC T cells specific for a tumour protein, and for inhibiting the development  
 CC of cancer especially prostate cancer. Compositions comprising the  
 CC polynucleotide and/or polypeptide are useful for stimulating an immune  
 CC response, and for treating cancer. The oligonucleotide is useful for  
 CC detecting cancer. The present sequence is a prostate specific  
 CC polypeptide of the invention.  
 XX  
 SQ Sequence 339 AA;

Query Match 100.0%; Score 192; DB 22; Length 339;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-18;  
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RRSRYKLTLMAYQVQVQNKEDAMIEHDVWRMEI 34  
 Db 185 RRSRYKLTLMAYQVQVQNKEDAMIEHDVWRMEI 218

RESULT 7  
 AAM78845  
 ID AAM78845 standard; Protein; 339 AA.

XX AAM78845;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human protein SEQ ID NO 1507.  
 XX  
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157190-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US04098.  
 XX  
 XX 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0606875.  
 PR 20-JUN-2000; 2000US-0598075.  
 PR 19-JUL-2000; 2000US-0620325.  
 PR 01-SEP-2000; 2000US-0654936.  
 PR 15-SEP-2000; 2000US-0663561.  
 PR 20-OCT-2000; 2000US-0693325.  
 PR 30-NOV-2000; 2000US-0728422.  
 XX  
 XX (HYSE-) HYSEO INC.

PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI

PI Xue AJ, Yang Y, Wejhrman T, Goodrich R,  
 XX  
 DR WPI: 2001-476283/51.  
 DR N-PsDB; AAK51978.  
 XX  
 PT Nucleic acids encoding polypeptides with cytokine-like activities,  
 PT useful in diagnosis and gene therapy -  
 XX  
 PS Claim 20; Page 3800-3801; 6221pp; English.

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activity/inhibit activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAM80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.  
 XX

SQ Sequence 339 AA;

Query Match 100.0%; Score 192; DB 22; Length 339;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-18;  
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RRSRYKLTLMAYQVQVQNKEDAMIEHDVWRMEI 34  
 Db 185 RRSRYKLTLMAYQVQVQNKEDAMIEHDVWRMEI 218

RESULT 8  
 AAM01282  
 ID AAM01282 standard; Protein; 339 AA.

XX AAM01282;  
 XX  
 DT 04-OCT-2001 (first entry)  
 XX  
 DE P789P amino acid sequence.  
 XX

KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;  
 KW cytostatic; gene therapy; metastasis.

OS Homo sapiens.  
 XX  
 PN WO200151633-A2.  
 XX  
 PD 19-JUL-2001.  
 XX  
 PF 16-JAN-2001; 2001WO-US01574.  
 XX  
 PR 14-JAN-2000; 2000US-0483672.  
 XX  
 XX (CORI-) CORIXA CORP.

PA Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;  
 PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skelky YAW,  
 PI Wang A, Meagher MJ;  
 XX  
 DR WPI: 2001-425873/45.  
 XX

PT New polynucleotide encoding a prostate-specific protein, for  
 PT diagnosing, monitoring and treating prostate cancer in a patient and  
 PT for use in vaccines -  
 XX  
 PS Claim 2; Page 510-512; 543pp; English.

XX

CC The present invention describes polynucleotide sequences (I) which encode  
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,  
CC and can be used in vaccine production and gene therapy. (I), (II),  
CC antibodies to (I), fusion proteins comprising (II), and isolated  
CC T cells prepared using (I) or (II) are used treat cancer in a patient.  
CC (I) and the antibodies are also used in the detection of cancer in a  
CC patient. The cancer that is diagnosed or treated is particularly  
CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or  
CC (I) can be used for monitoring the progression of cancer in a patient.  
CC (I) and (II) can also be used to improve diagnostic and therapeutic  
CC methods for prostate cancer. They can indicate the level of metastasis  
CC as well as the prostate volume. AAH9357 to AAH9394 and AAH0115 to  
CC AAH0138 represent polynucleotide and amino acid sequences used in the  
CC exemplification of the present invention.

SO Sequence 339 AA;

Query Match 100.0%; Score 192; DB 22; Length 339;  
Best Local Similarity 100.0%; Pred. No. 5.1e-18;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRSYRYKLNMAYQVOONKEDAMIEHDVWRMEI 34  
DB 185 RRSYRYKLNMAYQVOONKEDAMIEHDVWRMEI 218

## RESULT 9

ABG61813  
ID ABG61813 standard; Protein; 339 AA.

AC ABG61813;

DT 15-AUG-2002 (first entry)

DE Prostate cancer-associated protein #14.

KW Prostate cancer; prostate tumour tissue; human; mammal; cytostatic.

OS Mammalia.

PN MO200230268-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US32045.

PR 13-OCT-2000; 2000US-0687576.

PR 08-DEC-2000; 2000US-0733288.

PR 24-JAN-2001; 2001US-2639579.

PR 16-MAR-2001; 2001US-2767919.

PR 16-MAR-2001; 2001US-2768889.

PR 06-APR-2001; 2001US-2819229.

PR 24-APR-2001; 2001US-2862149.

PR 30-APR-2001; 2001US-0847046.

PR 04-MAY-2001; 2001US-2885899.

XX (EOSB-) EOS BIOTECHNOLOGY INC.

XX Gish KC, Mack DH, Wilson KE, Afar D, Hevez F,

XX WPI; 2002-471335/50.

XX N-PSDB; ABK92128.

XX Detecting a prostate cancer-associated transcript in a cell in a

XX PT patient, useful for diagnosing prostate cancer (PC) or screening

XX PT modulators of PC, by determining if prostate cancer-associated genes

XX are expressed in a prostate tissue -

XX Claim 27, Page 312; 436pp; English.

XX The present invention relates to methods of detecting a prostate

XX cancer-associated transcript in a cell from a patient. The method

CC comprises contacting a biological sample from the patient with  
CC prostate cancer-associated polynucleotides (designated PC genes) that  
CC selectively hybridize to a sequence that is at least 80% identical  
CC to them. The prostate cancer-associated polynucleotide sequences  
CC are differentially expressed in prostate tumour tissue or in  
CC prostate cancer and are derived from the tissues of various  
CC organisms such as humans or other mammals (e.g. mice, sheep and dogs).  
CC The methods of the invention are useful for diagnosing and treating  
CC prostate cancer in mammals. The prostate cancer-associated genes are  
CC useful for diagnosing or treating prostate cancer, as well as for  
CC identifying modulators of prostate cancer or agents that inhibit  
CC prostate cancer. The nucleic acid sequences are particularly useful  
CC in gene therapy, as a vaccine or in antisense applications.

SO Sequence 339 AA;

Query Match 100.0%; Score 192; DB 23; Length 339;  
Best Local Similarity 100.0%; Pred. No. 5.1e-18;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRSYRYKLNMAYQVOONKEDAMIEHDVWRMEI 34  
DB 185 RRSYRYKLNMAYQVOONKEDAMIEHDVWRMEI 218

## RESULT 10

ABB95387  
ID ABB95387 standard; Protein; 339 AA.

AC ABB95387;

DT 19-JUL-2002 (first entry)

DE Human P789P protein SEQ ID NO 879.

KW Human; cancer; prostate cancer; vaccine; cytostatic; immunostimulant;

OS Gene therapy.

PN Homo sapiens.

PD US2002022248-A1.

PF 21-FEB-2002.

PR 12-JAN-2001; 2001US-0759143.

PR 25-FEB-1997; 97US-0806099.

PR 01-AUG-1997; 97US-0904804.

PR 09-FEB-1998; 98US-0020956.

PR 25-FEB-1998; 98US-0030607.

PR 14-JUL-1998; 98US-0115453.

PR 23-SEP-1998; 98US-0159812.

PR 15-JAN-1999; 99US-0232149.

PR 09-APR-1999; 99US-0288946.

PR 13-JUL-1999; 99US-0352616.

PR 12-NOV-1999; 99US-0439313.

PR 18-NOV-1999; 99US-0443686.

PR 14-JAN-2000; 2000US-0483672.

PR 27-MAR-2000; 2000US-0536857.

PR 09-MAY-2000; 2000US-0568100.

PR 13-MAY-2000; 2000US-0570737.

PR 27-JUN-2000; 2000US-0605783.

PR 10-AUG-2000; 2000US-0636215.

PR 29-AUG-2000; 2000US-0651236.

PR 06-SEP-2000; 2000US-0657279.

PR 02-OCT-2000; 2000US-0679426.

PR 10-OCT-2000; 2000US-0685166.

XX (XUJ/) XU J.

XX (DILL/) DILLON D C.

XX (MITC/) MITCHAM J L.



XX	(HARL./	HARLOCKER S L.
PA	(JIANG/	JIANG Y.
PA	(KALOS/)	KALOS M D.
PA	(FRANG/)	FANGER G R.
PA	(RETT/)	RETTER M W.
PA	(STOL/)	STOLK J A.
PA	(DAYC/)	DAY C H.
PA	(VEDV/)	VEDVICK T S.
PA	(CART/)	CARTER D.
PA	(LISX/)	LI S X.
PA	(WANG/)	WANG A.
PA	(SKEI/)	SKEIKY Y A W.
PA	(HEPL/)	HEPLER W T.
XX	(HENND/)	HENDERSON R A.
XX	Xu J,	Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI	Fanger GB, Retter MM, Stolk JA, Day CH, Vedvick TS, Carter D;	
PI	Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;	
XX	WPI; 2002-255649/30.	
DR	New prostate-specific polynucleotides for diagnosing and treating	
PT	prostate diseases, in particular prostate cancer, and as markers for the	
PT	progression of cancer	
PS	Claim 2; SEQ ID NO 879; 87pp; English.	
XX	The present invention provides prostate-specific coding sequences and	
CC	their encoded proteins. These can be used in the diagnosis and treatment	
CC	of cancers, particularly prostate cancer. The present sequence is a	
CC	protein described in the invention.	
XX	Sequence 339 AA;	
XX	Sequence 339 AA;	

Query Match		100.0%	Score 192;	DB 23;	Length 339;
Best Local Similarity		100.0%	Pred. No. 5.1e-18;		
Matches 34;	Conservative	0;	Mismatches	0;	Gaps 0;
Indels		0;			
QY	1 RRSYRYKLLNMAVQQVOONKEADAWIEHDWRMEI	34			
Dd	185 RRSYRYKLLNMAVQQVOONKEADAWIEHDWRMEI	218			
RESULT 11					
ID	AAE02780				
XX	AAE02780 standard; Protein; 375 AA.				
XX	AAE02780;				
XX					
XX	06-AUG-2001 (first entry)				
DE	Human six transmembrane epithelial antigen of prostate (STEAP)-1 protein.				
XX					
XX	Human; cytosolic; antiproliferative; vaccine; gene therapy;				
KM	six transmembrane epithelial antigen of the prostate-1; STEAP-1;				
KW	chromosome 7p22.3; cancer; prostate; colon; bladder; lung; ovarian;				
pancreatic.					
XX					
XX	Homo sapiens.				
OS					
XX					
FH	Key	Location/Qualifiers			
FT	Region	14..28			
FT	/label= Immunogenic_peptide #1				
FT	Domain	70..91			
FT	/label= Transmembrane_domain #1				
FT	Region	86..94			
FT	/label= HLA-A2_binding_peptide #2				
FT	Domain	114..136			
FT	/label= Transmembrane_domain #2				
FT	Region	158..166			
FT	/label= HLA-A2_binding_peptide #5				
FT	Domain	163..184			
FT	/label= Transmembrane_domain #3				

```

FT FT Region /label= HLA-A2_binding_peptide #1
FT FT Domain 165..173
FT FT Domain 219..241
FT FT Domain /label= Transmembrane_domain #4
FT FT Domain 254..276
FT FT Region /label= Transmembrane_domain #5
FT FT Region 262..270
FT FT Domain /label= HLA-A2_binding_peptide #3
FT FT Domain 292..313
FT FT Region /label= Transmembrane_domain #6
FT FT Region 302..310
FT FT Misc-difference /label= HLA-A2_binding_peptide #4
FT FT 339..340
FT FT /note= "Encoded by TTGTGAGAT"
XX XX WO200140276-A2.
XX XX PD
XX XX PD 07-JUN-2001.
XX XX PR 06-DEC-2000; 2000WO-US33040.
XX XX PR 06-DEC-1999; 99US-0455486.
XX XX PA (UROC-) UROGENESYS INC.
XX XX P1 Afar DEH, Hubert RS, Raitano AB, Saffran DC, Faris M,
XX XX P1 Jakobovits A;
XX XX DR WPI; 2001-367804/38.
XX XX DR N-PSDB; AAD07067.
XX XX PT New STEAP (six transmembrane epithelial antigen of the prostate)
XX XX PT proteins, expressed in human cancers, useful for detecting and treating
XX XX PT cancer -
XX XX PS Example 2: Fig 1A-1B; 187pp; English.
XX XX CC The present sequence is human six transmembrane epithelial antigen of
XX XX CC the prostate (STEAP)-1 protein of clone 10. STEAP is a member of cell
XX XX CC surface serpentine transmembrane antigens. STEAP-1 gene is located on
XX XX CC chromosome 7p22.3 and is used in gene therapy. Inhibiting the development
XX XX CC or progression of a cancer (eg. prostate, colon, bladder, lung, ovarian
XX XX CC and pancreatic) expressing STEAP or inhibiting growth or killing cells
XX XX CC expressing STEAP in a patient, comprises administering a vaccine
XX XX CC composition to the patient. Treating a patient with a cancer that
XX XX CC expresses STEAP, or inhibiting growth or killing cells expressing STEAP,
XX XX CC comprises administering to the patient a vector encoding single chain
XX XX CC monoclonal antibody that comprises the variable domains of the heavy and
XX XX CC light chains of the monoclonal antibody that specifically binds to STEAP,
XX XX CC such that the vector delivers the single chain monoclonal antibody coding
XX XX CC sequence to the cancer cells and the encoded single chain monoclonal
XX XX CC Note: The present sequence is also shown in sequence listing of the
XX XX CC specification, but it lacks amino acid residues at its N-terminal end.
XX XX SQ Sequence 375 AA;
XX XX Query Match 100.0%; Score 192; DB 22; Length 375;
XX XX Best local similarity 100.0%; Pred. No. 5,7e-18;
XX XX Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX XX QY 1 RRSRYRKLINWAYQVOONKEDAWIEHDVWMEI 34
XX XX | |||||
XX XX Db 185 RRSRYRKLINWAYQVOONKEDAWIEHDVWMEI 218
XX XX |||||
XX XX RESULT 12
XX XX AAMW6309
XX XX ID AAMW6309 standard; Protein; 339 AA.
XX XX AC AAMW6309;
XX XX DT 01-MAR-1999 (first entry)

```

DE Kidney injury associated molecule HM018 protein.  
XX  
XX Kidney injury associated molecule; kidney injury related molecule;  
KW KIM; tissue growth promotion; regeneration; renal condition;  
XX acute renal failure; acute nephritis; tumour.  
XX  
XX Rattus sp.  
OS  
PN MO9853071.A1.  
PN 26-NOV-1998.  
PD  
PF 22-MAY-1998; 98MO-US10547.  
XX  
XX 23-MAY-1997; 97US-0047491.  
PR 23-MAY-1997; 97US-0047490.  
XX  
PA (BIOJ ) BIOGEN INC.  
PI Gate RL, Hession CA, Sanicola-Nadel M, Wei H;  
XX WPI; 1999-045312/04.  
DR N-PSDB; AAV80586.  
XX  
PT Kidney injury-associated molecule, KIM, polypeptides - upregulated  
PT in injured or regenerating tissues, useful to promote tissue growth  
PT and regeneration, especially to treat renal conditions  
XX  
PS Claim 17; Page 57-58; 213pp; English.  
XX  
CC The present sequence represents a kidney injury associated molecule  
CC (KIM) protein. KIM proteins can be administered therapeutically  
CC by expressing KIM encoding polynucleotides, to promote growth and/or  
CC survival of damaged tissue (e.g. renal tissue), since the KIM proteins  
CC are upregulated in injured or regenerating (especially renal) tissues.  
CC KIM fusion proteins, conjugates, antibodies and vectors can also be used  
CC therapeutically, e.g. these or the KIM proteins may be included with an  
CC acceptable carrier in pharmaceutical compositions, useful for therapy/  
CC prophylaxis of conditions associated with dysfunction/dysregulation of  
CC KIM genes or proteins, especially renal diseases or impairments of renal  
CC function in humans (e.g. acute renal failure, acute nephritis). The  
CC polynucleotides can be used to produce antisense sequences which, when  
CC internalised into cells, can disrupt expression of a cellular KIM gene,  
CC also useful in therapy (e.g. to block the growth of tumours dependent on  
CC KIM for growth) or compositions. The proteins and polynucleotides are  
CC useful diagnostically e.g. to detect and quantify renal injury/disease  
CC (indicative of increased risk, or presence of, renal injury or impaired  
CC function), or abnormal responses to tissue injury (indicative of  
CC increased risk, or presence of, an autoimmune response or abnormal  
CC tissue growth arising from/affecting renal tissue). The proteins can  
CC also be used to locate KIM-producing cells (especially specific loci,  
CC e.g. tissue masses abnormally producing/expressing KIM such as tumours  
CC arising from/affecting renal tissue), by contacting cells with an  
CC imageable KIM-binding reagent and imaging reagent accumulation.  
XX  
SQ Sequence 339 AA;  
XX  
Query Match 94.8%; Score 182; DB 20; Length 339;  
Best Local Similarity 91.2%; Pred. No. 1,2e-16;  
Matches 31; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RRSYRYKLLNMAVYQOVQNKEDAMIEHDVYRMEI 34  
|||||  
Db 185 RRSYRYKLLNMAVYQOVQSKEDAMVYHDVYRMEI 218  
|||||  
RESULT 13  
ID AAM79829  
XX AAM79829 standard; Protein; 374 AA.  
XX  
XX AAM79829;

```

DT      06-NOV-2001 (first entry)
XX
DE      Human protein SEQ ID NO 3475.
XX
KW      Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KM      vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KV      tissue growth factor; immunomodulatory; cancer; leukaemia;
XX      nervous system disorder; arthritis; inflammation.
OS      Homo sapiens.
PN      WO200157190-A2.
PD      09-AUG-2001.
XX
PF      05-FEB-2001; 2001WO-US04098.
XX
PR      03-FEB-2000; 2000US-0496914.
PR      27-APR-2000; 2000US-0560875.
PR      20-JUN-2000; 2000US-0598075.
PR      19-JUL-2000; 2000US-0620325.
PR      01-SEP-2000; 2000US-0654936.
PR      15-SEP-2000; 2000US-0663561.
PR      20-OCT-2000; 2000US-0693325.
PR      30-NOV-2000; 2000US-0728422.
XX
PA      (HYSE-) HYSEQ INC.
XX
PI      Tang YT, Liu C, Dymnac RT, Aundt V, Zhou P, Xu C, Cao Y, Ma Y;
PI      Zhao QJ, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI      Xue AJ, Yang Y, Wejhrman T, Goodrich R;
DR      WPI: 2001-476283/51.
XX
DR      N-PBDB; AAK52962.
XX
PT      Nucleic acids encoding polypeptides with cytokine-like activities,
PT      useful in diagnosis and gene therapy -
XX
PS      Claim 20; Page 350; 6221pp; English.
XX
CC      The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC      encoded polypeptides (AAM78323-AAH80302) that exhibit activity elating to
CC      cytokine, cell proliferation or cell differentiation or which may induce
CC      production of other cytokines in other cell populations. The
CC      polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC      peptide therapy. The polypeptides have various cytokine-like activities,
CC      e.g. stem cell growth factor activity, haematopoiesis regulating
CC      activity, tissue growth factor activity, immunomodulatory activity and
CC      activin/inhibin activity and may be useful in the diagnosis and/or
CC      treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC      inflammation.
CC      Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC      (AAM80020) are omitted as the relevant pages from the sequence listing
CC      were missing at the time of publication.
XX
SQ      Sequence    374 AA;
XX
Query Match          93.2%; Score 179; DB 22; Length 374;
Best Local Similarity 97.1%; Pred. No. 3,3e-16;
Matches   33; Conservative   0; Mismatches   1; Indels   0; Gaps   0;
QY      1 RRSYYRYKLIMAYQQVOONKEPAMTEHVWRMEI 34
ID      AAB49483
ID      AAB49483 standard; Protein; 488 AA.
XX
AC      AAB49483;
XX
DT      08-MAR-2001 (first entry)

```

```

XX Rat p-HYDE.
DE
XX Rat; p-HYDE; cytostatic; gene therapy; apoptosis; leukemia; prostate;
XX tumour suppressor gene; DNA repair; cancer; melanoma; lymphoma;
KM colorectal; pancreatic; breast; brain; gastric carcinoma.
XX
OS Rattus sp.
XX WO200071564-A2.
XX
XX 30-NOV-2000.
XX
XX 01-MAY-2000; 2000WO-US11456.
XX
XX 29-APR-1999; 99US-0131607.
PR 29-APR-1999; 99US-0302457.
PR 26-NOV-1999; 99US-0499817.
XX
XX (UYTE-) UNIV TENNESSEE RES CORP.
XX
XX Steiner MS, Wang C, Rinaldy A, Menon R;
PI WPI; 2001-032016/04.
XX
XX N-PSDB; AAC89167.
XX
XX New isolated nucleic acid encoding a mammalian p-Hyde protein of the
PT p-Hyde family is useful for treating cancer, e.g. prostate cancer -
XX
XX Claim 31; Page 20; 171pp; English.
XX
XX The present protein is rat p-HYDE. p-HYDE induces susceptibility of a
XX cancer cell to cell death. The p-HYDE gene is associated with the
XX regression of tumour growth in vivo, the induction to susceptibility to
XX apoptosis caused by UV or chemotherapy induced DNA damage and prevention
XX of DNA repair with the upregulation of apoptosis as the result of UV
XX damage and the failure to repair DNA. The present sequence may be used to
XX treat cancer, preferably melanoma, lymphoma, leukaemia, prostate,
XX colorectal, pancreatic, breast, brain or gastric carcinoma.
XX
XX Sequence 488 AA;
SQ
XX
XX Query Match 56.8%; Score 109; DB 22; Length 488;
XX Best Local Similarity 58.8%; Pred. No. 1.5e-06;
XX Matches 20; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
XX
XX 1 RRSYRYKLNMAYQOQONKEDAMIEHDVWRMEI 34
XX |||:||||:||||:||||:||||:||||:
DB 326 RRSRYDLVNLAVKQVLANKSRMLWEEVEWRMEI 359
XX
XX RESULT 15
XX ABB83365
XX ID ABB83365 standard; Protein; 526 AA.
XX
XX ABB83365;
XX
XX 02-SEP-2002 (first entry)
XX
XX Murine Tumour Suppressor Activated Pathway 6, TSAP6.
XX
XX Neuroprotective; cytostatic; TSAP6; apoptosis; cancer; neurodegeneration;
KM Tumour Suppressor Activated Pathway 6; TSAP6 binding protein; tumour;
XX murine.
XX
XX Mus musculus.
XX
XX WO200252274-A2.
XX
XX 04-JUL-2002.
XX
XX 24-DEC-2001; 2001WO-FR04188.
XX

```

```

PR 26-DEC-2000; 2000FR-0017027.
PR 18-SEP-2001; 2001WO-FR02896.
XX
XX (MOLE-) MOLECULAR ENGINES LAB.
XX
XX Amson R, Telerman A, Passer B;
XX
XX WPI; 2002-508914/54.
XX
XX N-PSDB; ABN85026.
XX
XX Identifying compounds that inhibit binding of TSAP6 to proteins for
PT treating cancer and neurodegeneration -
XX
XX Disclosure; Page 75-77; 79pp; French.
XX
XX The present invention relates to a method for identifying a compound
XX (I) that inhibits binding of TSAP (Tumour Suppressor Activated Pathway) 6
XX to a TSAP6 binding protein. The present sequence is murine TSAP6, which
XX was used in the method of the invention. (I) are useful for modulating
XX tumour reversal and/or apoptosis for treating cancer or
XX neurodegeneration.
XX
XX Sequence 526 AA;
SQ
XX
XX Query Match 56.8%; Score 109; DB 23; Length 526;
XX Best Local Similarity 58.8%; Pred. No. 1.6e-06;
XX Matches 20; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
XX
XX 1 RRSYRYKLNMAYQOQONKEDAMIEHDVWRMEI 34
XX |||:||||:||||:||||:||||:||||:
DB 364 RRSRYDLVNLAVKQVLANKSRMLWEEVEWRMEI 397
XX
XX Search completed: March 26, 2003, 16:48:15
XX Job time : 67.0233 secs

```



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 16:43:45 ; Search time 23.0233 Seconds  
(without alignments)  
196.889 Million cell updates/sec

Title: US-10-010-667a-19

Perfect score: 116  
Sequence: 1 REVHPLATSHQOYFYKIPILV 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virius:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	116	100.0	264	4 Q9Y6U5	Q9Y6U5 homo sapien
2	108	93.1	338	6 Q9GJ50	Q9GJ50 sus scrofa
3	94	81.0	339	11 Q9CWR7	Q9CWR7 mus musculu
4	94	81.0	339	11 Q924Z2	Q924Z2 mus musculu
5	94	81.0	339	11 Q924J9	Q924J9 mus musculu
6	48	41.4	172	2 Q938L7	Q938L7 streptococc
7	48	41.4	388	16 Q9A0P6	Q9A0P6 streptococc
8	48	41.4	456	4 Q8TDP3	Q8TDP3 homo sapien
9	48	41.4	487	4 Q8TF03	Q8TF03 homo sapien
10	48	41.4	488	4 Q9NVB5	Q9NVB5 homo sapien
11	48	41.4	929	12 Q9QAP4	Q9QAP4 cervid herp
12	47	40.5	233	10 Q9SHA6	Q9SHA6 arabidopsis
13	47	40.5	427	16 Q8R872	Q8R872 thermomast
14	47	40.5	431	17 Q9HWR3	Q9HWR3 halobacteri
15	47	40.5	469	4 Q9C0F0	Q9C0F0 homo sapien
16	47	40.5	565	3 Q04600	Q04600 saccharomyc

17	47	40.5	568	4 Q96MU3	Q96MU3 homo sapien
18	47	40.5	675	3 Q9UES5	Q9UES5 schizosacch
19	47	40.5	763	10 Q9FU78	Q9FU78 oryza sativ
20	47	40.5	763	10 Q9LW12	Q9LW12 oryza sativ
21	47	40.5	763	10 Q9LW12	Q9LW12 oryza sativ
22	47	40.5	806	10 Q9LW12	Q9LW12 oryza sativ
23	47	40.5	912	12 Q8QV01	Q8QV01 buffalo her
24	47	40.5	947	12 Q9QAP7	Q9QAP7 bovine herp
25	46.5	40.1	437	2 Q93TY7	Q93TY7 shigella fl
26	46	39.7	169	10 Q93865	Q93865 glycine max
27	46	39.7	240	2 Q8VVM7	Q8VVM7 vibrio chol
28	46	39.7	296	16 Q8Y8R2	Q8Y8R2 listeria mo
29	46	39.7	425	16 Q92E57	Q92E57 listeria in
30	46	39.7	425	16 Q8Y9D5	Q8Y9D5 listeria in
31	46	39.7	430	16 Q8X9P2	Q8X9P2 bacillus ha
32	46	39.7	473	11 Q99JG6	Q99JG6 mus musculu
33	46	39.7	488	11 Q99P41	Q99P41 rattus norv
34	46	39.7	509	11 Q99JG5	Q99JG5 mus musculu
35	46	39.7	514	11 Q924Z1	Q924Z1 mus musculu
36	46	39.7	735	16 Q82D01	Q82D01 yersinia pe
37	46	39.7	931	12 Q9QAP6	Q9QAP6 rangiferine
38	45	38.8	151	5 Q9WSA2	Q9WSA2 drosophila
39	45	38.8	304	10 Q9SVB7	Q9SVB7 arabidopsis
40	45	38.8	428	10 Q9M2P5	Q9M2P5 arabidopsis
41	45	38.8	436	10 Q8RXF2	Q8RXF2 arabidopsis
42	45	38.8	470	2 Q9AKD0	Q9AKD0 rickettsia
43	45	38.8	607	16 Q8Y172	Q8Y172 bruceella me
44	45	38.8	919	12 Q9QAP5	Q9QAP5 caprine her
45	44.5	38.4	308	17 Q8THL2	Q8THL2 mechanosarc

## ALIGNMENTS

### RESULT 1

Q9Y6U5 PRELIMINARY; PRT; 264 AA.  
ID Q9Y6U5  
AC Q9Y6U5;  
DT 01-NOV-1999 (T-REMBLrel. 12, Created)  
DT 01-NOV-1999 (T-REMBLrel. 12, last sequence update)  
DT 01-DEC-2001 (T-REMBLrel. 19, last annotation update)  
DE WGS: H. RG087E15.1 protein (Fragment).  
GN WGS: H. RG087E15.1  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99063792; PubMed=9847074;  
RA Sulston J.E., Waterston R.;  
RT "Toward a complete human genome sequence.";  
RN [2]  
RP GENOME RES. 8:1097-1108(1998).  
RA Strong C., Layman D., Graves T., Strommatt C.;  
RT "The sequence of Homo sapiens BAC clone CTR-87E15.";  
RN [3]  
RP Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RA SEQUENCE FROM N.A.  
RA Waterston R.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC005061; AAD43182.1; -;  
FT NON\_TER 1 1  
FT NON\_TER 264 264  
SQ SEQUENCE 264 AA; 31203 MW; 35C9483003557E72 CRC64;

Query Match 100.0%; Score 116; DB 4; Length 264;  
Best Local Similarity 100.0%; Pred. No. 3.1e-11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REVHPLATSHQOYFYKIPILV 22  
|||||

DB 102 REVHPLATSHOQFYKIPILV 123

RESULT 2

Q9GL50 PRELIMINARY; PRT; 338 AA.

AC Q9GL50;

DT 01-MAR-2001 (TEMBLrel. 16, Created)

DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)

DE Six transmembrane endothelial antigen of PAEC.

OS Sus scrofa (Pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.

OC NCBI\_TaxID=9823;

OX [1]

RP SEQUENCE FROM N.A.

RA Nagaoka T., Boulevarde G., Coupel S., Coulon F., Tesson L., Heblan J.-M., Soullion J.-P., Charreau B.;

RT "Differential gene expression in endothelial cells during TNF-alpha-RT and LPS-mediated activation."

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF319659; AAG33868.1; -

KW Transmembrane.

SQ SEQUENCE 338 AA; 39918 MW; ED490E86E067A32B CRC64;

Query Match 93.1%; Score 108; DB 6; Length 338;

Best Local Similarity 90.9%; Pred. No. 8.8e-10; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 REVHPLATSHOQFYKIPILV 22

Db 91 REVHPLATSHOQFYKIPILV 112

RESULT 3

Q9CMR7 PRELIMINARY; PRT; 339 AA.

AC Q9CMR7;

DT 01-JUN-2001 (TEMBLrel. 17, Created)

DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)

DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)

DE 2410007B19RIK protein.

GN STEAP OR 2410007B19RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI\_TaxID=10090;

OX [1]

RP SEQUENCE FROM N.A.

RA STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai T., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I., Saito T., Ozaki Y., Gotohori T., Bono H., Kasukawa T., Saito R., Kadoya K., Matsuda H.A., Ashburner M., Batelov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H., Kuehl P., Lewis S., Maleno Y., Nikaido I., Pesole G., Quackenbush J., Schirini L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barah G., Blake J., Boffelli D., Bojunga N., Carinci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Guelinckx S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P., Norde P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohzuki S., Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection."

RL Nature 409:685-690(2001).

DR EMBL; AK010437; BAB26938.1; -

DR MGD; MGI:1917608; Steap.

SQ SEQUENCE 339 AA; 39264 MW; 3F7AB9C7520F0968 CRC64;

Query Match 81.0%; Score 94; DB 11; Length 339;

Best Local Similarity 77.3%; Pred. No. 1.9e-07; Indels 0; Gaps 0;

Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 REVHPLATSHOQFYKIPILV 22

Db 92 REVHPLATSHOQFYKIPILV 113

RESULT 4

Q924Z2 PRELIMINARY; PRT; 339 AA.

AC Q924Z2;

DT 01-DEC-2001 (TEMBLrel. 19, Created)

DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)

DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)

DE Dudulin.

GN 1010001D01RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI\_TaxID=10090;

OX [1]

RP SEQUENCE FROM N.A.

RA Serru V., Manivet P., Lambdin D., Vaudourdolle M., Kellermann O., Loric S.;

RT "Prostate and non-prostate expression of dudulin, the mouse ortholog of human STEAP."

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY029584; AAK50537.1; -

DR MGD; MGI:1915678; 1010001D01RIK.

SQ SEQUENCE 339 AA; 39105 MW; 32A2C29F2E333BD0 CRC64;

Query Match 81.0%; Score 94; DB 11; Length 339;

Best Local Similarity 77.3%; Pred. No. 1.9e-07; Indels 0; Gaps 0;

Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 REVHPLATSHOQFYKIPILV 22

Db 92 REVHPLATSHOQFYKIPILV 113

RESULT 5

Q924U9 PRELIMINARY; PRT; 339 AA.

AC Q924U9;

DT 01-DEC-2001 (TEMBLrel. 19, Created)

DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)

DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)

DE Six-transmembrane epithelial antigen of the prostate.

GN STEAP.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI\_TaxID=10090;

OX [1]

RP SEQUENCE FROM N.A.

RA STRAIN=C57BL/6J; PubMed=11479226;

RX MEDLINE=21371909; PubMed=11479226;

RA Yang D., Holt G.E., Velders M.P., Kwon B.D., Kaat W.M.;

RT "Murine six-transmembrane epithelial antigen of the prostate, prostate stem cell antigen, and prostate-specific membrane antigen: prostate-specific cell-surface antigens highly expressed in prostate cancer of transgenic adenocarcinoma mouse prostate mice."

RL Cancer Res. 61:5857-5860(2001).

DR EMBL; AF297098; AAK83126.1; -

DR MGD; MGI:1917608; Steap.

KW Transmembrane.

SQ SEQUENCE 339 AA; 39318 MW; 4B26A71FF559E84F CRC64;

```

Query Match      81.0%;  Score 94;  DB 11;  Length 339;
Best Local Similarity  77.3%;
Matches 17;  Conservative  3;  Mismatches  2;  Indels  0;  Gaps  0

```

```

RESULT 6
093BL7
ID      093BL7          PRELIMINARY;          PRT;          172 AA.
AC      093BL7;
DT      01-DEC-2001 (TREMBLrel. 19, Created)
DT      01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT      01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE      Hypothetical 19.5 kDa protein.
OS      Streptococcus pyogenes
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales
OC      Streptococcaceae; Streptococcus.
OX      NCBI_TaxID=114;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=NH1;
RA      Ikebe T., Wada A., Inegaki Y., Sugama K., Tanaka D., Suzuki R.,
RA      Katsukawa C., Fujinaga Y., Aoe Y., Watanabe H.;
RL      "Complete sequence of temperate phage phiNH1.1."
RL      Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AY050245; ALA15065.1; -.
DR      InterPro: IPR003115; PfamB:
DR      Pfam: PF02195; ParBc; 1.
KM      Hypothetical protein.
SQ      SEQUENCE 172 AA; 19484 MW; AD8165655EDAEEFC CRC64;

```

	Query Match	41.4%	Score 48:	DB 2:	length 172:
	Best Local Similarity	38.1%	Pred. No. 4.4:		
	Matches	8;	Conservative	8;	Mismatches 5; Indels 0; Gaps 0
Qy	2 EVIHLPLATSHQQYFYKPIRLV 22	:   :	:	:	:
Dh	23 EAVGPVAASIKKEFGFKVPILV 43	:   :	:	:	:

RESULT 7			
Q9A0P6			
ID	Q9A0P6	PRELIMINARY;	PRT; 388 AA.
AC	Q9A0P6;		
DT	01-JUN-2001 (TrEMBLrel. 17, Created)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Hypothetical phage associated protein.		
GN	SPY0679.		
OS	Streptococcus pyogenes.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;		
OC	Streptococcaceae; Streptococcus.		
OX	NCBI_TaxID=1314;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=SF370 / ATCC 700294 / Serotype M1;		
RX	MEDLINE=21192694; PubMed=11296296;		
RA	Ferretti U.T., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,		
RA	Primeaux C., Seate S., Shvortov A.N., Kenton S., Lai H.S., Lin S.P.,		
RA	Qian Y., Jia H.G., Najaf F.Z., Ren Q., Zhu H., Song L., White J.,		
RA	Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;		
RT	"Complete genome sequence of an M1 strain of <i>Streptococcus pyogenes</i> ."		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:4658-4663(2001).		
DR	EMBL; A8006521; AAK3638.1; -		
DR	InterPro; IPR001091; CN4_Mettransf.		
DR	InterPro; IPR002295; D21N6_Mtase.		
DR	InterPro; IPR002941; N6/N4_Mtase.		
DR	InterPro; IPR002052; N6_Mtase.		
DR	InterPro; IPR003115; PaBic.		
DR	PIfam; PF01535; N6_N4_Mtase; 1.		

DR PFam: PF02195; ParBc\_1.  
DR PRINTS: PR00506; D1N6MFERASE.  
DR PRINTS: PR00508; S214MFERASE.  
DR SMART: SM00470; ParBc\_1.  
DR PROSITE: PS00092; N6\_MTASE; UNKNOWN\_1.  
KW Hypothetical protein, complete genome.  
SQ SEQUENCE 388 AA; 43650 MW; 12587B620606EB7 CRC64;

		Query Match	41.4%	Score 48;	DB 16;	length 388;
		Best Local Similarity	38.1%;	Pred. No. 11;		
		Matches	8;	Conservative	8;	Mismatches
					5;	Indels
					0;	Gaps
Q7	2	EVIHPLATSHOOQFYKPIIV	22			
		: :   : : :   :   :				.
Db	23	EAVGPVASEIKFEGFKPIIV	43			

RESULT 8	
Q8TDP3	
ID	Q8TDP3
AC	Q8TDP3;
DT	01-JUN-2002 (TREMBlrel. 21, Created)
DT	01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE	PHYDE II.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxId=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Wang C., Allay J.A., Steiner M.S.;
RT	"Second human member of PHYDE family, Human PHYDE II.";
RL	Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
DR	EMBL, AP262322, AAM08128.1, --
SQ	SEQUENCE 456 AA; 50430 MW; CSF7C7008D55251E CRC64;

Query Match	41.4%	Score 48	DB 4	Length 456
Best Local Similarity	36.4%	Pred. No. 13		
Matches	8	Conservative	5	Mismatches 9
				Indels 0
				Gaps 0
QY	1	REVTHPLATSHQQYRYKIPILY	22	
		: : : : : : : : : : :		
Db	233	RDVLQPVYQESONKFEKLPVSV	254	

QY	1	REV HPLATSHOQVFXPKIPILV	22
DB	233	RDVLOPFIWQESONKFFKXLPVSU	254

RESULT 10  
 Q9NVB5 PRELIMINARY; PRT; 488 AA.  
 AC Q9NVB5; 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE cDNA FLJ10829 fis, clone NT2RP4001138 (Dudulin 2).  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Iwagaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Aotaka S., Yoshikawa Y.,  
 RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,  
 RA Nakamura Y., Nagahari K., Masuko Y., Sasaki N.;  
 RT "NEDO human cDNA sequencing project."  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Seru V., Manivet P., Lenoir C., Eschwege P., Lambin D.,  
 RA Vaudourdolle M., Kellermann O., Loric S.;  
 RT "Dudulin 2, a new tumor antigen expressed in various human tumors."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK001691; BAA91839.1; -;  
 DR EMBL; AY029585; AAK50538.1; -;  
 DR InterPro; IPR003006; IG\_MHC.  
 DR PROSITE; PS00290; IG\_MHC; UNKNOWN 1.  
 SO SEQUENCE 488 AA; 54616 MW; BC0BCA48335AAD6 CRC64;

Query Match 41.4%; Score 48; DB 4; Length 488;  
 Best Local Similarity 36.4%; Pred. No. 14;  
 Matches 8; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 REVHPLATSHOOYFYKIPILV 22  
 Db 233 RVLIQPIYQESQNKFKLPVSV 254

RESULT 11  
 Q9QAP4 PRELIMINARY; PRT; 929 AA.  
 AC Q9QAP4; 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Glycoprotein B.  
 OS Cervid herpesvirus 1.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae.  
 NCBI\_TaxID=79891;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=BAKFBSHIRE 82;  
 RX MEDLINE=99221732; PubMed=10203465;  
 RA Ros C., Belak S.;  
 RT "Studies of genetic relationships between bovine, caprine, cervine,  
 RT and rangeliferine alphaherpesviruses and improved molecular methods for  
 RT virus detection and identification."  
 RL J. Clin. Microbiol. 37:1247-1253(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=BAKFBSHIRE 82;  
 RC Ros C., Belak S.;  
 RT "Characterization of the glycoprotein B gene from ruminant  
 RT alphaherpesviruses."  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF078729; AAD46115.2; -;  
 SO SEQUENCE 929 AA; 100755 MW; BC0E5692230BAB CRC64;

Query Match 41.4%; Score 48; DB 12; Length 929;  
 Best Local Similarity 46.7%; Pred. No. 27;  
 Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 REVHPLATSHOOYFYKIP 15  
 Db 655 RELVEPCANHKRYF 669

RESULT 12  
 Q9SHA6 PRELIMINARY; PRT; 233 AA.  
 AC Q9SHA6; 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE At2g37810 protein.  
 GN AT2G37810.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., Vanden S.E., Umayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 RT thaliana."  
 RL Nature 402:761-768(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Lin X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC007661; AAD32790.1; -;  
 DR InterPro; IPR000345; CYC2\_heme\_bind.  
 DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.  
 SO SEQUENCE 233 AA; 26250 MW; 2266F40BD1CEC910 CRC64;

Query Match 40.5%; Score 47; DB 10; Length 233;  
 Best Local Similarity 43.8%; Pred. No. 8;  
 Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 IHPATSHOOYFYKIP 19  
 Db 105 VHPCTQHCHVSYP 120

RESULT 13  
 Q8R872 PRELIMINARY; PRT; 427 AA.  
 AC Q8R872; 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE O-acetylthiosulfinate sulfinylase.  
 GN MET17 OR TIE2151.  
 OS Thermococcuszoochloae tengcongensis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 OC Thermococcobacteriales; Thermococcobacteriaceae; Thermococcobacter.  
 NCBI\_TaxID=119072;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MB4T / JCM11007;



RX MEDLINE=21992816; PubMed=11997336;  
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;  
 RT "A complete sequence of T. tengcongensis genome.";  
 RL Genome Res. 12:689-700(2002).  
 DR EMBL; AE013162; AAM25314.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 427 AA; 46642 MW; 0A4B1F2796CEE4AB CRC64;

Query Match 40.5%; Score 47; DB 16; Length 427;  
 Best Local Similarity 72.7%; Pred. No. 17;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VIHPATSHQO 13  
 DB 376 VIHPASTTHQO 386

## RESULT 14

09HMR3 PRELIMINARY; PRT; 431 AA.  
 AC 09HMR3;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)  
 DE O-acetyl homoserine.  
 GN HAL OR VNG2421G.  
 OS Halobacterium sp. (strain NRC-1).  
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 OC Halobacteriaceae; Halobacterium.  
 OX NCBI\_TaxId=64091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahatras G.G., Bergquist B., Pan M.,  
 RA Shukla H.D., Laskey S.R., Baliga N.S., Thorson V., Shrogha J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Geo Y.A.,  
 RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angewine C.M., Dale H.,  
 RA Isenbarger T.A., Peck R.F., Pohnschroder M., Spudich J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;  
 RT "Genome sequence of Halobacterium species NRC-1.";  
 RL proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 DR EMBL; AE005123; AAG20508.1; -.  
 DR HSSP; P00935; ICSI.  
 DR InterPro; IPR000277; Cys\_Met\_Meta\_PP.  
 DR Pfam; PF01053; Cys\_Met\_Meta\_PP; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 431 AA; 45768 MW; 67C6C5064A2DB9B3 CRC64;

Query Match 40.5%; Score 47; DB 17; Length 431;  
 Best Local Similarity 72.7%; Pred. No. 17;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VIHPATSHQO 13  
 DB 378 VIHPASTTHQO 388

## RESULT 15

09COP0 PRELIMINARY; PRT; 469 AA.  
 AC 09COP0;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)  
 DT 01-OCT-2001 (TREMBLrel. 18, last annotation update)  
 DE KIAA1713 protein (Fragment).  
 GN KIAA1713.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21082932; PubMed=11214970;  
 RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XIX.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 7:347-355(2000).  
 DR EMBL; AB051500; BAB21804.1; -.  
 DR InterPro; IPR003980; H3\_histamine\_rec.  
 DR InterPro; IPR002965; P\_Fich\_extensn.  
 DR PRINTS; PRO1471; HISTAMTHER3R.  
 DR PRINTS; PRO1217; PRICEXTENSN.  
 FT NON TER 1  
 SQ SEQUENCE 469 AA; 50649 MW; 235F3AFA4B627525 CRC64;

Query Match 40.5%; Score 47; DB 4; Length 469;  
 Best Local Similarity 58.3%; Pred. No. 19;  
 Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 9 TSHQOYFYKIP 20  
 DB 146 TSHQOYFYQMPV 157

Search completed: March 26, 2003, 16:49:44  
 Job time : 25.0233 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 16:44:15 ; Search time 12.5349 Seconds  
(without alignments)  
168.726 Million cell updates/sec

Title: US-10-010-667a-19

Perfect score: 116  
Sequence: 1 REVHPLATSHQOYFYKIPILV 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1: \*  
2: p1r2: \*  
3: p1r3: \*  
4: p1r4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52.5	45.3	416	2	E64140
2	49	42.2	177	2	G84273
3	47.5	40.9	320	2	S73186
4	47	40.5	233	2	D84797
5	47	40.5	431	2	H84392
6	47	40.5	565	2	S52682
7	47	40.5	675	2	T39727
8	46	39.7	169	2	T07623
9	46	39.7	296	2	AD1178
10	46	39.7	425	2	AD1149
11	46	39.7	425	2	AD1508
12	46	39.7	428	2	T08576
13	46	39.7	430	2	C83975
14	46	39.7	735	2	AD0341
15	46	39.7	737	2	C70770
16	46	39.7	1839	1	OYBYK
17	45	38.8	304	2	T09357
18	45	38.8	428	2	T46025
19	45	38.8	607	2	AG3123
20	44.5	38.4	136	2	S72508
21	44.5	38.4	1176	2	A27826
22	44	37.9	163	2	C91092
23	44	37.9	163	2	G85937
24	44	37.9	472	2	G81293
25	44	37.9	771	2	S51421
26	44	37.9	928	1	VGEBBC
27	44	37.9	932	1	VGEBBC
28	44	37.9	969	2	B87336
29	43.5	37.5	326	2	S56534

30	43.5	37.5	326	2	D91287	hypothetical prote
31	43.5	37.5	326	2	G86128	hypothetical prote
32	43	37.1	284	2	S71227	extensin I - Arabi
33	43	37.1	295	2	F83356	hypothetical prote
34	43	37.1	336	2	F86319	hypothetical prote
35	43	37.1	337	1	S14523	aspartate-semialde
36	43	37.1	338	2	B82118	probable aspartate
37	43	37.1	373	2	B96798	extensin [imported
38	43	37.1	379	2	C86873	transcription regu
39	43	37.1	384	2	B72324	oxaloacetate decar
40	43	37.1	385	2	T16447	hypothetical prote
41	43	37.1	426	2	A86634	O-acetylhomoserine
42	43	37.1	430	2	D72324	O-acetylhomoserine
43	43	37.1	434	2	AH3397	cysteine synthase
44	43	37.1	442	2	T44655	O-acetylhomoserine
45	43	37.1	485	2	T37550	hypothetical colle

ALIGNMENTS

RESULT 1  
E64140  
hypothetical protein H10021 - Haemophilus influenzae (strain Rd KW20)  
C/Species: Haemophilus influenzae  
C/Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 08-Oct-1999  
C/Accession: E64140  
R/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kiknes, E.F.; Kerlavage, A.  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Meidman, J.  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.  
A/Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A/Reference number: A64000; MUID:95350630; PMID:7542800  
A/Accession: E64140  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-416 <TIGR>  
A/Cross-references: GB:U32687; GB:IA2023; NID:G1572955; PIDN:AAC21699.1; PID:G1572965; TJ  
A/Note: best homolog was a hypothetical protein from Klebsiella pneumoniae  
C/Genetics:  
A/Start codon: GTG

Query Match 45.3% Score 52.5; DB 2; Length 416;  
Best local similarity 50.0%; Pred. No. 1.5;  
Matches 10; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 1 REVHPLATSHQOYFYKIPILV 20  
Db 34 KEIIRPLETCHAEVAV-LPI 52

RESULT 2  
G84273  
peptide methionine sulfoxide reductase [imported] - Halobacterium sp. NRC-1  
C/Species: Halobacterium sp. NRC-1  
C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C/Accession: G84273  
R/My, M.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Laaky, S.  
; Leitbauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor  
Jung, K.H.; Alam, M.; Frettas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A/Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Edhardt, H.; Lowe, T.M.; Lie  
A/Title: Genome sequence of Halobacterium species NRC-1.  
A/Reference number: AB4160; MUID:20504463; PMID:11016950  
A/Accession: G84273  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-177 <STO>  
A/Cross-references: GB:AE004437; NID:G10580713; PIDN:AA619555.1; GSPDB:GN00138  
C/Genetics:  
A/Gene: mraA  
C/Superfamily: peptide methionine sulfoxide reductase

Query Match 42.2%; Score 49; DB 2; Length 177;  
 Best Local Similarity 50.0%; Pred. No. 2.1;  
 Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EVIHLPLATSHOOYFYKIP 19  
 |||:|||||  
 Db 135 ETFFPALEKQNYEKNP 152

RESULT 3  
 573186  
 cytochrome f - red alga (Porphyra purpurea) chloroplast  
 C:/Species: chloroplast Porphyra purpurea  
 C:/Date: 19-Mar-1997 #sequence\_revision 09-May-1997 #text\_change 11-Jun-1999  
 C:/Accession: S73186

R/Reich, M.; Munnolland, J.  
 Plant Mol. Biol. Rep. 13, 333-335, 1995  
 A/Title: Complete nucleotide sequence of the Porphyra purpurea chloroplast genome.  
 A/Reference number: S73108  
 A/Accession: S73186  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-320 <REI>  
 A/Cross-references: EMBL:U38804; NID:G1276552; PIDN:AAC08151.1; PID:G1276731  
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995  
 C/Genetics:

A/Genes: petA  
 A/Genome: chloroplast  
 C/Superfamily: cytochrome f  
 C/Keywords: chloroplast; heme; thylakoid

Query Match 40.9%; Score 47.5; DB 2; Length 320;  
 Best Local Similarity 40.7%; Pred. No. 7.1;  
 Matches 11; Conservative 4; Mismatches 7; Indels 5; Gaps 1;

QY 1 REVILHP-----ATSHOOYFYKIPILV 22  
 |||:|||||  
 Db 160 REIFPILSPDPAKDQAHFKYPIIV 166

RESULT 4  
 D84797  
 hypothetical protein At2g37810 [imported] - Arabidopsis thaliana  
 C:/Species: Arabidopsis thaliana (mouse-ear cress)  
 C:/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:/Accession: D84797

R/Lin, X.; Kaul, S.; Roundley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
 Gues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A/Reference number: A84420; MUID:2003487; PMID:10617197

A/Accession: D84797  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-233 <STO>  
 A/Cross-references: GB:AE002093; NID:G4895203; PIDN:AAD32790.1; GSPDB:GN00139  
 C/Genetics:

A/Genes: At2g37810  
 A/Map position: 2

Query Match 40.5%; Score 47; DB 2; Length 233;  
 Best Local Similarity 43.8%; Pred. No. 6;  
 Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 IHPPLATSHOOYFYKIP 19  
 |||:|||||  
 Db 105 VHPCTQHGHVSVIP 120

RESULT 5  
 H84392

O-acetyl homoserine [imported] - Halobacterium sp. NRC-1  
 C:/Species: Halobacterium sp. NRC-1  
 C:/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
 C:/Accession: H84392

R/Hy, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Laeky, S.;  
 Leitchauer, B.; Keller, K.; Cruz, R.; Dawson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor  
 Jung, K.H.; Alam, M.; Freilias, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A/Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Edhardt, H.; Lowe, T.M.; Li

A/Title: Genome sequence of Halobacterium species NRC-1.  
 A/Reference number: A84160; MUID:20504483; PMID:11016950  
 A/Accession: H84392  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-431 <STO>  
 A/Cross-references: GB:AE004437; NID:G10581828; PIDN:AG20508.1; GSPDB:GN00138  
 C/Genetics:

A/Genes: hal  
 C/Superfamily: O-succinylhomoserine (thiol)-lyase

Query Match 40.5%; Score 47; DB 2; Length 431;  
 Best Local Similarity 72.7%; Pred. No. 12;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VHPPLATSHOO 13  
 |||:|||||  
 Db 378 VHPASTTHOO 368

RESULT 6  
 S52682  
 hypothetical protein YDR117C - yeast (Saccharomyces cerevisiae)  
 N/Alternate names: hypothetical protein YD9727.12c.  
 C:/Species: Saccharomyces cerevisiae  
 C:/Date: 19-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 19-Apr-2002  
 C/Accession: S52682

R/Murphy, L.; Shore, L.; Harris, D.  
 submitted to the EMBL Data Library, March 1995  
 A/Reference number: S52671  
 A/Accession: S52682

A/Molecule type: DNA  
 A/Residues: 1-565 <MUR>  
 A/Cross-references: EMBL:Z48758; NID:G747879; PID:G747891; GSPDB:GN00004; MIPS:YDR117C  
 C/Genetics:

A/Genes: MIPS:YDR117C  
 A/Cross-references: SCD:S0002524  
 A/Map position: 4R

Query Match 40.5%; Score 47; DB 2; Length 565;  
 Best Local Similarity 47.1%; Pred. No. 16;  
 Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 EVIHLPLATSHOOYFYKIP 18  
 |||:|||||  
 Db 439 EILHPLLTNNTEFYQI 455

RESULT 7  
 T39727  
 nucleoporin homolog - fission yeast (Schizosaccharomyces pombe)  
 C:/Species: Schizosaccharomyces pombe  
 C:/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 07-Dec-1999  
 C/Accession: T39727

R/Mood, V.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.  
 submitted to the EMBL Data Library, August 1999  
 A/Reference number: Z21874  
 A/Accession: T39727  
 A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA  
 A/Residues: 1-675 <MOO>  
 A/Cross-references: EMBL:AL109846; PIDN:CA852802.1; GSPDB:GN00067; SPDB:SPBC17G9.04C  
 A/Experimental source: strain 972h-; cosmid c17G9

C/Genetics:

A:Gene: SPDB:SPEC17G9.04c  
A:Map position: 2

Query Match 40.5%; Score 47; DB 2; Length 675;  
Best Local Similarity 57.1%; Pred. No. 20;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 VIHPLATSHQOYFY 16  
DB 566 IISPAITHQOYFY 579

RESULT 8  
T07623

extensin homolog HRGP2 - soybean (fragment)  
N:Alternate names: hydroxyproline-rich glycoprotein HRGP2

C:Species: Glycine max (soybean)  
C:Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 11-May-2000

C:Accession: T07623  
R:Hong, J.C.; Cheong, Y.H.; Nagao, R.T.; Bahk, J.D.; Cho, M.J.; Key, J.L.  
Plant Physiol. 104, 793-796, 1994

A:Title: Isolation and characterization of three soybean extensin cDNAs.  
A:Reference number: Z16058; MUID:94211912; PMID:8159793

A:Accession: T07623  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA  
A:Residues: 1-169 <HON>

A:Cross-references: EMBL:122030; NID:g347454; PIDN:AAA33971.1; PID:g347455

A:Experimental source: strain Wayne; seedling  
C:Genetics:  
A:Gene: HRGP2  
C:Superfamily: hydroxyproline-rich glycoprotein  
C:Keywords: glycoprotein; hydroxyproline

Query Match 39.7%; Score 46; DB 2; Length 169;  
Best Local Similarity 57.1%; Pred. No. 6;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 6 PLATSHQOYFYKIP 19  
DB 92 PSPTSHPPYTKSP 105

RESULT 9  
A11178

transcription regulator homolog lmo0833 [imported] - *Listeria monocytogenes* (strain EGD-  
C:Species: *Listeria monocytogenes*  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

C:Accession: A11178  
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, R.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;  
D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkac, G.; Madueno, E.; Maltournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,

A:Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: A11178  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-296 <GLA>

A:Cross-references: GB:NC 003210; PIDN:CAC98911.1; PID:g1641022; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo0833

Query Match 39.7%; Score 46; DB 2; Length 296;  
Best Local Similarity 41.2%; Pred. No. 11;  
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 2 EVIHLPLATSHQOYFYKT 18  
DB 222 EILHELSTHDCAFYRI 238

RESULT 10  
AD1149

O-acetylhomoserine sulphydrylase homolog lmo0595 [imported] - *Listeria monocytogenes* (str  
C:Species: *Listeria monocytogenes*  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

C:Accession: AD1149  
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, R.;  
Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;  
D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkac, G.; Madueno, E.; Maltournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,

A:Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AD1149  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-425 <GLA>

A:Cross-references: GB:NC 003210; PIDN:CAC98674.1; PID:g16409971; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo0595

C:Superfamily: O-succinylhomoserine (thiol)-lyase  
Query Match 39.7%; Score 46; DB 2; Length 425;  
Best Local Similarity 63.6%; Pred. No. 17;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 VIHPLATSHQO 13  
DB 376 IHPASTTHQO 386

RESULT 11  
AD1508

O-acetylhomoserine sulphydrylase homolog lmo604 [imported] - *Listeria innocua* (strain C)  
C:Species: *Listeria innocua*  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

C:Accession: AD1508  
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, R.;  
Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;  
D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkac, G.; Madueno, E.; Maltournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,

A:Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AD1508  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-425 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC95836.1; PID:g16413044; GSPDB:GN00178  
A:Experimental source: strain C11p11262  
C:Genetics:  
A:Gene: lmo604

C:Superfamily: O-succinylhomoserine (thiol)-lyase  
Query Match 39.7%; Score 46; DB 2; Length 425;  
Best Local Similarity 63.6%; Pred. No. 17;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 VIHPLATSHQO 13  
DB 376 IHPASTTHQO 386

RESULT 12  
T08576

phenylalanine-tRNA ligase (EC 6.1.1.20) beta chain T22F8.180 [similarity] - *Arabidopsis t*  
C:Species: *Arabidopsis thaliana* (mouse-ear cross)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 03-Jun-2002

C/Accession: T08576  
 R/Bevan, M.; Zimmermann, W.; Grueneisen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; May  
 submitted to the Protein Sequence Database, May 1999  
 A/Reference number: Z16442  
 A/Accession: T08576  
 A/Molecule type: DNA  
 A/Residues: 1-428 <BEV>  
 A/Cross-references: EMBL:AL050351, GSPDB:GN0062, ATSP:T22F8.180  
 A/Experimental source: cultivar Columbia; BAC clone T22F8  
 C/Genetics:  
 A/Map position: 4  
 A/Insertions: 49/3; 91/3; 121/3; 140/3; 179/2; 196/3; 213/1; 241/2; 273/3; 306/1; 330/1; 35  
 C/Superfamily: yeast cytosolic phenylalanine-tRNA ligase beta chain  
 C/Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 39.7%; Score 46; DB 2; Length 428;  
 Best Local Similarity 46.7%; Pred. No. 17;  
 Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 5 HPLATSHOQFYKIP 19  
 DB 200 HPARDSHDTFLKVP 214

RESULT 13  
 C83975  
 O-cetylhomoserine sulphydrolase BH2603 [imported] - Bacillus halodurans (strain C-125)  
 C/Species: Bacillus halodurans  
 C/Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C/Accession: C83975  
 R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A/Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A/Reference number: A83650; MUID:20512582; PMID:11058132  
 A/Accession: C83975  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-430 <STO>  
 A/Cross-references: GB:RP001516; GB:BA000004; NID:g10175192; PIDN:BAB06322.1; GSPDB:GN00  
 A/Experimental source: strain C-125  
 C/Genetics:  
 A/Gene: BH2603  
 C/Superfamily: O-succinylhomoserine (thiol)-lyase

Query Match 39.7%; Score 46; DB 2; Length 430;  
 Best Local Similarity 63.6%; Pred. No. 17;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 VHPLATSHOQ 13  
 DB 379 IHHPASTTHQ 389

RESULT 14  
 AD0341  
 probable membrane protein YPO2801 [imported] - Yersinia pestis (strain CO92)  
 C/Species: Yersinia pestis  
 C/Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
 C/Accession: AD0341  
 R/Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tildall, R.W.; Holden, M.T.G.; Prentice, M.B.;  
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
 H. M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,  
 Nature 413, 523-527, 2001  
 A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A/Reference number: AB0001; MUID:21470413; PMID:11586360  
 A/Accession: AD0341  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-735 <KUR>  
 A/Cross-references: GB:AL590842; PIDN:CAC93035.1; PID:ig15980773; GSPDB:GN00175  
 C/Genetics:  
 A/Gene: YPO2801

Query Match 39.7%; Score 46; DB 2; Length 735;  
 Best Local Similarity 50.0%; Pred. No. 32;  
 Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 VHPLATSHOQFYKIPILV 22  
 DB 529 LHWLAPDNOQLFFFLPIL 548

RESULT 15  
 C70770  
 hypothetical protein RV1327c - Mycobacterium tuberculosis (strain H37Rv)  
 C/Species: Mycobacterium tuberculosis  
 C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C/Accession: C70770  
 R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A/Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A/Reference number: A70500; MUID:98295987; PMID:9634230  
 A/Accession: C70770  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-737 <COL>  
 A/Cross-references: GB:Z73902; GB:AL123456; NID:g3261576; PIDN:CAA98091.1; PID:e245021.1  
 A/Experimental source: strain H37Rv  
 C/Genetics:  
 A/Gene: RV1327c

Query Match 39.7%; Score 46; DB 2; Length 737;  
 Best Local Similarity 46.7%; Pred. No. 32;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 HPLATSHOQFYKIP 19  
 DB 389 HPWAREHRCWFELP 403

Search completed: March 26, 2003, 16:50:38  
 Job time: 14.5349 secs

GenCore version 5.1.4\_p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 16:38:14 ; Search time 42.7209 Seconds  
(without alignments)  
68.620 Million cell updates/sec

Title: US-10-010-667a-19

Perfect score: 116  
Sequence: 1 REVHPLATSHQQYFYKIPILV 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

```
A.Geneseq_101002.*
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	116	100.0	22	21	AA58198 Human STRAP-1 pept
2	116	100.0	22	22	AAE02786 Extracellular loop
3	116	100.0	104	22	ABB40417 Peptide #7923 enco
4	116	100.0	104	22	AAW73944 Human bone marrow
5	116	100.0	104	23	ABG43832 Human peptide enco
6	116	100.0	339	21	AA58194 Human STRAP-1 prot
7	116	100.0	339	22	AAU69927 Human prostate cDN
8	116	100.0	339	22	AAW78845 Human protein SEQ
9	116	100.0	339	22	AAW01282 P789P amino acid s
10	116	100.0	339	23	ABG61813 Prostate cancer-as

11	116	100.0	339	23	ABB95387 Human P789P protei
12	116	100.0	374	22	AAW79829 Human protein SEQ
13	116	100.0	375	22	AAE02780 Human six transmem
14	97	83.6	95	20	AAV11840 Human 5' EST seque
15	92	79.3	109	20	AAV12304 Human 5' EST seque
16	89	76.7	339	20	AAW86309 Kidney injury asso
17	68	58.6	26	22	AAE02789 Extracellular loop
18	68	58.6	419	22	AAU10189 Human ORF2 of Six-
19	68	58.6	450	22	AAE02841 Human STEAP-2 prot
20	68	58.6	454	22	AAU10188 Human ORF2 of Six-
21	68	58.6	454	22	AAE02781 Human six transmem
22	68	58.6	454	22	ABG61933 Prostate cancer-as
23	68	58.6	454	23	AAU80190 Human PUMPn prote
24	68	58.6	454	23	AAU76538 Tumour-associated
25	68	58.6	490	22	AAU10187 Human Six-Transmem
26	68	58.6	576	22	ABG12306 Novel human diagno
27	68	58.6	1273	22	ABG00113 Novel human diagno
28	49	42.2	128	22	AAO10072 Human polypeptide
29	48	41.4	237	22	AAU04565 Human G-protein co
30	48	41.4	388	22	ABP29233 Streptococcus poly
31	48	41.4	456	22	ABP49482 Human p-HYDE 40.
32	48	41.4	487	22	ABP49481 Human p-HYDE. Hom
33	48	41.4	488	22	AAU10220 Human Six-Transmem
34	48	41.4	488	22	AAU04564 Human G-protein co
35	48	41.4	488	22	AAW85775 Human drug metabol
36	48	41.4	488	22	ABP3224 Human protein sequ
37	48	41.4	488	23	ABB83366 Human Tumour Supp
38	46	39.7	195	23	ABP38703 Staphylococcus epi
39	46	39.7	296	23	ABB47546 Listeria monocytog
40	46	39.7	425	23	ABB48993 Listeria monocytog
41	46	39.7	488	22	AAW49483 Rat p-HYDE. Ratu
42	46	39.7	526	23	ABB83365 Novel human diagno
43	45	38.8	146	22	ABG08176 Drosophila melanog
44	45	38.8	151	22	ABG69782
45	44.5	38.4	1176	22	ABBS9692

#### ALIGNMENTS

RESULT 1	AA58198	AA58198 standard; peptide; 22 AA.
ID	AA58198	
XX	AA58198;	
DT	14-MAR-2000	(first entry)
XX		
XX	Human STRAP-1 peptide, corresponding to STRAP-1 extracellular region 1.	
DE		
XX	Serpentine transmembrane antigen of the prostate: STRAP-1; prostate;	
KW	transmembrane domain; type IIIa membrane protein; expression; cancer;	
KW	prostate cancer; bladder cancer; colon cancer; pancreatic cancer;	
KW	ovarian cancer; tumour antigen; immunisation; immune response;	
KW	cellular; humoral; anticancer vaccine; antibody; detection; diagnosis;	
KW	prognosis; monitoring; susceptibility; therapeutic inhibitor;	
KW	drug targeting; recombinant protein.	
XX		
OS	Synthetic.	
OS	Homo sapiens.	
XX		
PN	MO9962941-A2.	
XX		
PD	09-DEC-1999.	
XX		
XX	01-JUN-1999;	99MO-US12157.
PF		
XX		
PR	01-JUN-1998;	98US-0087520.
PR	30-JUN-1998;	98US-0091183.
XX		
XX	(UROC-) UROGENESYS INC.	
PA	(AFAR/) AFAR D E.	
PA	(HUBE/) HUBERT R S.	

PA (LEON/) LEONG K.  
 PA (RAIT/) RAITANO A B.  
 PA (SAFF/) SAFFRAN D C.  
 XX  
 PI Afar DE, Hubert RS, Leong K, Raitano AB, Saffran DC;  
 XX WPI; 2000-072832/06.  
 XX  
 PT Novel proteins useful as diagnostic markers and therapeutic targets,  
 PT particularly for prostatic cancer -  
 XX  
 PS Disclosure; Page 22; 83pp; English.  
 XX  
 XX Sequences AAY58198-Y58200 represent synthetic peptides that correspond  
 CC to the extracellular regions of STRAP-1 (serpentine transmembrane  
 CC antigen of the prostate, AAY58194). These peptides were used to raise  
 CC monoclonal anti-STRAP-1 antibodies. STRAP-1 is the prototype  
 CC member of the STRAP family of proteins (AAY58194-Y58197) which  
 CC exhibit a high degree of structural conservation, but which show  
 CC no significant structural homology to known human proteins. The STRAP-1  
 CC gene has been localised to chromosome 7p22. STRAP-1 is thought to be a  
 CC type IIa membrane protein and is expressed predominantly in prostate  
 CC cells in normal human tissues. Structurally, STRAP-1 is a 339 amino  
 CC acid protein characterised by six transmembrane domains and  
 CC intracellular N- and C-termini, suggesting that it folds in a  
 CC "serpentine" manner into three extracellular and two intracellular loops.  
 CC STRAP-1 mRNA and protein expression is maintained at high levels and  
 CC throughout all stages of prostate cancer. STRAP-1 mRNA and/or protein is  
 CC also overexpressed in certain other cancers, including bladder, colon,  
 CC pancreatic and ovarian cancer. The function of the STRAP proteins is not  
 CC known. They may be ion channels (from the presence of six transmembrane  
 CC domains, a feature which is shared by certain ion channels) or  
 CC gap-junction proteins (from immunohistochemical staining). STRAP-1 and  
 CC STRAP-2 are cell-surface tumour antigens. Immunisation with a STRAP  
 CC protein induces cellular and humoral immune responses against  
 CC STRAP-expressing cells. STRAP proteins may be used to identify  
 CC specific-binding agents, to produce anticancer vaccines and to generate  
 CC specific antibodies. The antibodies may be used for detection, prognosis,  
 CC and monitoring of cancers (or susceptibility to cancer), as therapeutic  
 CC inhibitors or to target therapeutic agents to their site of action. STRAP  
 CC nucleic acids may be used for recombinant protein production, as  
 CC diagnostic and prognostic reagents, for identifying STRAP-expressing  
 CC cells for screening inhibitors of STRAP expression and for therapeutic  
 CC modulation/inhibition of STRAP expression. Since high levels of STRAP  
 CC proteins are exposed on the cell surface, they are easily targeted by  
 CC systematically administered agents, and because they are expressed mainly  
 CC on prostatic epithelial cells, agents targeted to them should have  
 CC minimal side effects on other tissues.  
 CC  
 XX  
 XX Sequence 22 AA;  
 SO  
 QY Query Match 100.0%; Score 116; DB 21; Length 22;  
 Db Best Local Similarity 100.0%; Pred. No. 7e-12;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 REVHPLATSHQYFYKIPILV 22  
 Db 1 REVHPLATSHQYFYKIPILV 22  
 XX  
 XX RESULT 2  
 AAE02786  
 ID AAE02786 standard; peptide; 22 AA.  
 XX  
 AC AAE02786;  
 XX  
 DT 06-AUG-2001 (first entry)  
 XX  
 DE Extracellular loop #1 of human STRAP-1, suitable for cloning into pfc.  
 XX  
 XX Human; cytostatic; antiproliferative; vaccine; gene therapy;  
 KW six transmembrane epithelial antigen of the prostate-1; STRAP-1;  
 KW chromosome 7p22.3; cancer; prostate; colon; bladder; pancreatic; lung;

KW ovarian; extracellular loop; serpentine transmembrane antigen.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200140276-A2.  
 XX  
 PD 07-JUN-2001.  
 XX  
 PF 06-DEC-2000; 2000MO-US33040.  
 XX  
 PR 06-DEC-1999; 99US-0455486.  
 XX  
 XX (UROC-) UROGENESYS INC.  
 PA  
 PI Afar DEH, Hubert RS, Raitano AB, Saffran DC, Mitchell SC, Farris M;  
 PI Jakobovits A;  
 XX WPI; 2001-367804/38.  
 DR  
 XX  
 XX New STRAP (six transmembrane epithelial antigen of the prostate)  
 PT proteins, expressed in human cancers, useful for detecting and treating  
 PT cancer -  
 XX  
 PS Example 19; Page 102; 187pp; English.  
 XX  
 XX The present invention relates to human six transmembrane epithelial  
 CC antigen of the prostate (STRAP) protein. STRAP is a member of cell  
 CC surface serpentine transmembrane antigens. STRAP gene is used in gene  
 CC therapy. Inhibiting the development or progression of a cancer (eg.  
 CC prostate, colon, bladder, lung, ovarian and pancreatic) expressing STRAP  
 CC or inhibiting growth or killing cells expressing STRAP in a patient,  
 CC comprises administering a vaccine composition to the patient. Treating  
 CC a patient with a cancer that expresses STRAP, or inhibiting growth or  
 CC killing cells expressing STRAP, comprises administering to the patient a  
 CC vector encoding single chain monoclonal antibody that comprises the  
 CC variable domains of the heavy and light chains of the monoclonal antibody  
 CC that specifically binds to STRAP, such that the vector delivers the  
 CC single chain monoclonal antibody coding sequence to the cancer cells and  
 CC the encoded single chain monoclonal antibody is expressed  
 CC intracellularly. The present sequence is extracellular loop of STRAP-1  
 CC suitable for cloning into pfc, which is used in the invention. STRAP-1  
 CC gene is located on chromosome 7p22.3.  
 CC  
 XX  
 XX Sequence 22 AA;  
 SO  
 QY Query Match 100.0%; Score 116; DB 22; Length 22;  
 Db Best Local Similarity 100.0%; Pred. No. 7e-12;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 REVHPLATSHQYFYKIPILV 22  
 Db 1 REVHPLATSHQYFYKIPILV 22  
 XX  
 XX RESULT 3  
 ABB40417  
 ID ABB40417 standard; Peptide; 104 AA.  
 XX  
 AC ABB40417;  
 XX  
 DT 04-FEB-2002 (first entry)  
 XX  
 DE Peptide #7923 encoded by human foetal liver single exon probe.  
 XX  
 XX Human; foetal liver; gene expression; single exon nucleic acid probe.  
 KW  
 OS Homo sapiens.  
 XX  
 PN WO200157277-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001MO-US00669.



```

XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX
XX Claim 27; SEQ ID NO 33052; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a peptide encoded by a single exon
XX nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pcc_sequences.
XX
XX Sequence 104 AA;
XX
XX Query Match 100.0%; Score 116; DB 22; Length 104;
XX Best Local Similarity 100.0%; Pred. No. 3.8e-11;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 REVYHPLATSHOOYFYKIPILV 22
XX |||||||||||||||||||
XX 59 REVYHPLATSHOOYFYKIPILV 80
XX
XX RESULT 4
XX AAM73944
XX ID AAM73944 standard; Protein; 104 AA.
XX
XX AAM73944;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 34250.
XX
XX Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; Leukaemia; myeloma.
XX
XX Homo sapiens.
XX
XX WO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00668.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX

```

```

XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 34250; 658pp + Sequence listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.
XX
XX Sequence 104 AA;
XX
XX Query Match 100.0%; Score 116; DB 22; Length 104;
XX Best Local Similarity 100.0%; Pred. No. 3.8e-11;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 REVYHPLATSHOOYFYKIPILV 22
XX |||||||||||||||||||
XX 59 REVYHPLATSHOOYFYKIPILV 80
XX
XX RESULT 5
XX ABG43832
XX ID ABG43832 standard; Peptide; 104 AA.
XX
XX ABG43832;
XX
XX 19-AUG-2002 (first entry)
XX
XX Human peptide encoded by genome-derived single exon probe SEQ ID 33497.
XX
XX Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Rudrak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease.
XX
XX Homo sapiens.
XX
XX WO200186003-A2.
XX
XX 15-NOV-2001.
XX
XX 30-JAN-2001; 2001WO-US00665.
XX
XX 04-FEB-2000; 2000US-180312P.
XX 26-MAY-2000; 2000US-207456P.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-234687P.
XX 27-SEP-2000; 2000US-236359P.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2002-114183/15.
XX
XX Spatially-addressable set of single exon nucleic acid probes, used to
XX measure gene expression in human lung samples -
XX
XX Claim 27; SEQ ID NO 33497; 634pp; English.
XX
XX

```



CC STRAP-expressing cells. STRAP proteins may be used to identify  
 CC specific-binding agents, to produce anticancer vaccines and to generate  
 CC specific antibodies. The antibodies may be used for detection, prognosis,  
 CC and monitoring of cancers (or susceptibility to cancer), as therapeutic  
 CC inhibitors or to target therapeutic agents to their site of action. STRAP  
 CC nucleic acids may be used for recombinant protein production, as  
 CC diagnostic and prognostic reagents, for identifying STRAP-expressing  
 CC cells for screening inhibitors of STRAP expression and for therapeutic  
 CC modulation/inhibition of STRAP expression. Since high levels of STRAP  
 CC proteins are exposed on the cell surface, they are easily targeted by  
 CC systemically administered agents, and because they are expressed mainly  
 CC on prostatic epithelial cells, agents targeted to them should have  
 CC minimal side effects on other tissues.

XX Sequence 339 AA;

Query Match 100.0%; Score 116; DB 21; Length 339;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-10;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REVHPLATSHQGYFYKPIPLV 22  
 DB 92 REVHPLATSHQGYFYKPIPLV 113

#### RESULT 7

ID AAU69927 standard; Protein; 339 AA.

AC AAU69927;

DT 30-JAN-2002 (first entry)

DE Human prostate cDNA encoded protein #72.

XX Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.

OS Homo sapiens.

PN WO200173032-A2.

XX 04-OCT-2001.

PF 27-MAR-2001; 2001WO-US09919.

XX 27-MAR-2000; 2000US-0536857.

PR 09-MAY-2000; 2000US-0568100.

PR 12-MAY-2000; 2000US-0570737.

PR 13-JUN-2000; 2000US-0593793.

PR 27-JUN-2000; 2000US-0605783.

PR 10-AUG-2000; 2000US-0636215.

PR 29-AUG-2000; 2000US-0651236.

PR 06-SEP-2000; 2000US-0657279.

PR 02-OCT-2000; 2000US-0679426.

PR 10-OCT-2000; 2000US-0685166.

XX (CORI-) CORIXA CORP.

XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;

PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;

PI Li SX, Wang A, Skeily YAM, Hepler WT, Henderson RA;

XX WPI; 2001-639232/73.

DR N-PSDB; AAS64160.

XX New human prostate-specific polypeptides and polynucleotides useful for

PT the diagnosis and treatment of cancer, especially prostate cancer -

XX Claim 2; Page 549; 579pp; English.

CC derived from them) and antigen-presenting cells expressing the  
 CC polypeptides. The antibodies are useful for detecting the presence of  
 CC cancer, especially prostate cancer. The polypeptides, polynucleotides and  
 CC the antigen-presenting cells are useful for stimulating and/or expanding  
 CC T cells specific for a tumour protein, and for inhibiting the development  
 CC of cancer especially prostate cancer. Compositions comprising the  
 CC polynucleotide and/or polypeptide are useful for stimulating an immune  
 CC response, and for treating cancer. The oligonucleotide is useful for  
 CC detecting cancer. The present sequence is a prostate specific  
 CC polypeptide of the invention.

XX Sequence 339 AA;

Query Match 100.0%; Score 116; DB 22; Length 339;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-10;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REVHPLATSHQGYFYKPIPLV 22  
 DB 92 REVHPLATSHQGYFYKPIPLV 113

#### RESULT 8

ID AAM78845 standard; Protein; 339 AA.

AC AAM78845;

DT 06-NOV-2001 (first entry)

DE Human protein SEQ ID NO 1507.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell; growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;

XX nervous system disorder; arthritis; inflammation.

OS Homo sapiens.

PN WO200157190-A2.

XX 09-AUG-2001.

PF 05-FEB-2001; 2001WO-US04098.

XX 03-FEB-2000; 2000US-0496814.

PR 27-APR-2000; 2000US-0560875.

PR 20-JUN-2000; 2000US-0598075.

PR 19-JUL-2000; 2000US-0620325.

PR 01-SEP-2000; 2000US-0654936.

PR 15-SEP-2000; 2000US-0663561.

PR 20-OCT-2000; 2000US-0693325.

PR 30-NOV-2000; 2000US-0728422.

XX (HYSE-) HYSEO INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;

PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI; 2001-476283/51.

DR N-PSDB; AAK51978.

XX Nucleic acids encoding polypeptides with cytokine-like activities,

PT useful in diagnosis and gene therapy -

XX Claim 20; Page 3800-3801; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the

CC encoded polypeptides (AAM78823-AAW80302) that exhibit activity elating to

CC cytokine, cell proliferation or cell differentiation and which may induce

CC production of other cytokines in other cell populations. The

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoietic regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC actinin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAM0020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.

XX  
 SQ Sequence 339 AA;

Query Match 100.0%; Score 116; DB 22; Length 339;  
 . Best Local Similarity 100.0%; Pred. No. 1.4e-10;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REVHPLATSHQOYFYKIPILV 22  
 DB 92 REVHPLATSHQOYFYKIPILV 113

RESULT 9  
 AAM01282

ID AAM01282 standard; Protein; 339 AA.

XX  
 AC AAM01282;

XX  
 DT 04-OCT-2001 (first entry)

XX  
 DE P789P amino acid sequence.

XX  
 KM Human; prostate cancer; prostate-specific; diagnosis; vaccine;  
 KM cytostatic; gene therapy; metastasis.

XX  
 OS Homo sapiens.

XX  
 PN WO200151633-A2.

XX  
 PD 19-JUL-2001.

XX  
 PF 16-JAN-2001; 2001WO-US01574.

XX  
 PR 14-JAN-2000; 2000US-0483672.

XX  
 PA (CORI-) CORIXA CORP.

XX  
 PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG,  
 PI Kaloje MD, Fanger GR, Day CH, Ketter MW, Scolk JA, Skelky YAM;  
 PI Wang A, Meagher MJ;

XX  
 DR WPI; 2001-425873/45.

XX  
 PT New polynucleotide encoding a prostate-specific protein, for  
 PT diagnosing, monitoring and treating prostate cancer in a patient and  
 PT for use in vaccines -

XX  
 PS Claim 2; Page 510-512; 543pp; English.

XX  
 CC The present invention describes polynucleotide sequences (I) which encode  
 CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,  
 CC and can be used in vaccine production and gene therapy. (I), (II),  
 CC antibodies to (II), fusion proteins comprising (II), and isolated  
 CC T cells prepared using (I) or (II) are used treat cancer in a patient.  
 CC (I) and the antibodies are also used in the detection of cancer in a  
 CC patient. The cancer that is diagnosed or treated is particularly  
 CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or  
 CC (I) can be used for monitoring the progression of cancer in a patient.  
 CC (I) and (II) can also be used to improve diagnostic and therapeutic  
 CC methods for prostate cancer. They can indicate the level of metastasis  
 CC as well as the prostate volume. AAK93357 to AAK93944 and AAM01115 to  
 CC AAM01318 represent polynucleotide and amino acid sequences used in the  
 CC exemplification of the present invention.

SQ Sequence 339 AA;

Query Match 100.0%; Score 116; DB 22; Length 339;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-10;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REVHPLATSHQOYFYKIPILV 22  
 DB 92 REVHPLATSHQOYFYKIPILV 113

RESULT 10  
 ABG61813

ID ABG61813 standard; Protein; 339 AA.

XX  
 AC ABG61813;

XX  
 DT 15-AUG-2002 (first entry)

XX  
 DE Prostate cancer-associated protein #14.

XX  
 KM Prostate cancer; prostate tumour tissue; human; mammal; cytostatic.

XX  
 OS Mammalia.

XX  
 PN WO200230268-A2.

XX  
 PD 18-APR-2002.

XX  
 PF 12-OCT-2001; 2001WO-US32045.

XX  
 PR 13-OCT-2000; 2000US-0687576.

XX  
 PR 08-DEC-2000; 2000US-0733388.

XX  
 PR 08-DEC-2000; 2000US-0733742.

XX  
 PR 24-JAN-2001; 2001US-263957P.

XX  
 PR 16-MAR-2001; 2001US-276791P.

XX  
 PR 16-MAR-2001; 2001US-276888P.

XX  
 PR 06-APR-2001; 2001US-281922P.

XX  
 PR 24-APR-2001; 2001US-286214P.

XX  
 PR 30-APR-2001; 2001US-0847046.

XX  
 PR 04-MAY-2001; 2001US-285589P.

XX  
 PA (EOSB-) EOS BIOTECHNOLOGY INC.

XX  
 PI Gish KC, Mack DH, Wilson KE, Afari D, Hevezi P;

XX  
 DR WPI; 2002-471335/50.

XX  
 DR N-FSDB; ABK92128.

XX  
 PT Detecting a prostate cancer-associated transcript in a cell in a  
 PT patient, useful for diagnosing prostate cancer (PC) or screening  
 PT modulators of PC, by determining if prostate cancer-associated genes  
 PT are expressed in a prostate tissue -

XX  
 PS Claim 27; Page 312; 436pp; English.

XX  
 CC The present invention relates to methods of detecting a prostate  
 CC cancer-associated transcript in a cell from a patient. The method  
 CC comprises contacting a biological sample from the patient with  
 CC prostate cancer-associated polynucleotides (designated PC genes) that  
 CC selectively hybridise to a sequence that is at least 80% identical  
 CC to them. The prostate cancer-associated polynucleotide sequences  
 CC are differentially expressed in prostate tumour tissue or in  
 CC prostate cancer and are derived from the tissues of various  
 CC organisms such as humans or other mammals (e.g. mice, sheep and dogs).

XX  
 CC The methods of the invention are useful for diagnosing and treating  
 CC prostate cancer in mammals. The prostate cancer-associated genes are  
 CC useful for diagnosing or treating prostate cancer, as well as for  
 CC identifying modulators of prostate cancer or agents that inhibit  
 CC prostate cancer. The nucleic acid sequences are particularly useful  
 CC in gene therapy, as a vaccine or in antisense applications.

XX  
 CC ABG61800-ABG61944 represent prostate cancer-associated proteins.

XX

SQ Sequence 339 AA;  
Query Match 100.0%; Score 116; DB 23; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1.4e-10;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 REVHPLATSHQQYFYKIPILV 22  
DB 92 REVHPLATSHQQYFYKIPILV 113  
RESULT 11  
ABBS5387  
ID ABB95387 standard; Protein; 339 AA.  
XX  
AC ABB95387;  
XX  
DT 19-JUL-2002 (first entry)  
XX  
DE Human P789P protein SEQ ID NO 879.  
XX  
KW Human; cancer; prostate cancer; vaccine; cytostatic; immunostimulant;  
KW gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN US200202248-A1.  
XX  
PD 21-FEB-2002.  
XX  
PF 12-JAN-2001; 2001US-0759143.  
XX  
PR 25-FEB-1997; 97US-0806099.  
PR 01-AUG-1997; 97US-0904804.  
PR 09-FEB-1998; 98US-0020956.  
PR 25-FEB-1998; 98US-0030607.  
PR 14-JUL-1998; 98US-0115453.  
PR 23-SEP-1998; 98US-0159812.  
PR 15-JAN-1999; 99US-0232149.  
PR 09-APR-1999; 99US-0288946.  
PR 13-JUL-1999; 99US-0352616.  
PR 12-NOV-1999; 99US-0439313.  
PR 18-NOV-1999; 99US-0443686.  
PR 14-JAN-2000; 2000US-0483672.  
PR 27-MAR-2000; 2000US-0536857.  
PR 09-MAY-2000; 2000US-0568100.  
PR 12-MAY-2000; 2000US-0570737.  
PR 13-JUN-2000; 2000US-0593793.  
PR 27-JUN-2000; 2000US-0605783.  
PR 10-AUG-2000; 2000US-0636215.  
PR 29-AUG-2000; 2000US-0651236.  
PR 06-SEP-2000; 2000US-0657279.  
PR 02-OCT-2000; 2000US-0679426.  
PR 10-OCT-2000; 2000US-0685166.  
XX  
XX (XUJ/) XU J.  
PA (DILL/) DILLON D C.  
PA (MITC/) MITCHAM J L.  
PA (HARL/) HARLOCKER S L.  
PA (JIAN/) JIANG Y.  
PA (KALO/) KALOS M D.  
PA (FANG/) FANGER G R.  
PA (RETT/) RETTER M W.  
PA (STOL/) STOLK J A.  
PA (DAYC/) DAY C H.  
PA (VEDV/) VEDVICK T S.  
PA (CART/) CARTER D.  
PA (LISX/) LI S X.  
PA (WANG/) WANG A.  
PA (SKEI/) SKEIKY Y A W.  
PA (HEPL/) HEPLER W T.  
PA (HEND/) HENDERSON R A.  
XX

PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;  
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;  
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;  
XX  
DR WPI; 2002-255649/30.  
XX  
PT New prostate-specific polynucleotides for diagnosing and treating  
PT diseases, in particular prostate cancer, and as markers for the  
PT progression of cancer -  
XX  
PS Claim 2; SEQ ID NO 879; 87pp; English.  
XX  
CC The present invention provides prostate-specific coding sequences and  
CC their encoded proteins. These can be used in the diagnosis and treatment  
CC of cancers, particularly prostate cancer. The present sequence is a  
CC protein described in the invention.  
XX  
SQ Sequence 339 AA;  
Query Match 100.0%; Score 116; DB 23; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1.4e-10;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 REVHPLATSHQQYFYKIPILV 22  
DB 92 REVHPLATSHQQYFYKIPILV 113  
RESULT 12  
AAM79829  
ID AAM79829 standard; Protein; 374 AA.  
XX  
AC AAM79829;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human protein SEQ ID NO 3475.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200157190-A2.  
XX  
PD 09-AUG-2001.  
XX  
PR 05-FEB-2001; 2001WO-US04098.  
XX  
PR 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX  
DR WPI; 2001-476283/51.  
XX  
DR N-PSDB; AAK52962.  
XX  
PT Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
XX  
PS Claim 20; Page 350; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activity/inhibit activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAW80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.  
 XX  
 SQ Sequence 374 AA;  
 Query Match 100.0%; Score 116; DB 22; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-10;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 REVHPLATSHQOYFYKIPILV 22  
 Db 127 REVHPLATSHQOYFYKIPILV 148  
 RESULT 13  
 AAE02780 ID AAE02780 standard; Protein; 375 AA.  
 XX  
 AC AAE02780;  
 XX  
 DT 06-AUG-2001 (first entry)  
 XX  
 DE Human s1x transmembrane epithelial antigen of prostate (STREP)-1 protein.  
 XX  
 KM Human; cytostatic; antiproliferative; vaccine; gene therapy;  
 KM s1x transmembrane epithelial antigen of the prostate-1; STREP-1;  
 KM chromosome 7p22.3; cancer; prostate; colon; bladder; lung; ovarian;  
 KM pancreatic.  
 KM  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT 14..28 /label= Immunogenic\_peptide #1  
 FT /label= Transmembrane\_domain #1  
 FT 86..94 /label= HLA-A2\_binding\_peptide #2  
 FT 114..136 /label= Transmembrane\_domain #2  
 FT 158..166 /label= Transmembrane\_domain #5  
 FT 163..184 /label= HLA-A2\_binding\_peptide #5  
 FT 165..173 /label= Transmembrane\_domain #3  
 FT /label= HLA-A2\_binding\_peptide #1  
 FT 219..241 /label= Transmembrane\_domain #4  
 FT 254..276 /label= Transmembrane\_domain #5  
 FT 263..270 /label= HLA-A2\_binding\_peptide #3  
 FT 292..313 /label= Transmembrane\_domain #6  
 FT 302..310 /label= Transmembrane\_domain #6  
 FT /label= HLA-A2\_binding\_peptide #4  
 FT Misc-difference 339..340 /note= "Encoded by TTGTAGAAAT"  
 FT  
 XX  
 PN WO200140276-A2.

XX  
 PD 07-JUN-2001.  
 XX  
 PF 06-DEC-2000; 2000WO-US33040.  
 XX  
 PR 06-DEC-1999; 99US-0455486.  
 XX  
 PA (UROC-) UROGENESYS INC.  
 XX  
 PI Afar DEH, Hubert RS, Raitano AB, Saffran DC, Mitchell SC, Faris M;  
 XX Jakobovits A;  
 XX WPI; 2001-367804/38.  
 DR N-PSDB; AAD07067.  
 XX  
 PT New STREP (s1x transmembrane epithelial antigen of the prostate)  
 PT protein, expressed in human cancers, useful for detecting and treating  
 PT cancer -  
 XX  
 PS Example 2; Fig 1A-1B; 187pp; English.  
 XX  
 CC The present sequence is human s1x transmembrane epithelial antigen of  
 CC the prostate (STREP)-1 protein of clone 10. STREP-1 gene is located on  
 CC surface serpinine transmembrane antigens. STREP-1 gene is located on  
 CC chromosome 7p22.3 and is used in gene therapy. Inhibiting the development  
 CC or progression of a cancer (eg. prostate, colon, bladder, lung, ovarian  
 CC and pancreatic) expressing STREP or inhibiting growth or killing cells  
 CC expressing STREP in a patient, comprises administering a vaccine  
 CC composition to the patient. Treating a patient with a cancer that  
 CC expresses STREP, or inhibiting growth or killing cells expressing STREP,  
 CC comprises administering to the patient a vector encoding single chain  
 CC monoclonal antibody that comprises the variable domains of the heavy and  
 CC light chains of the monoclonal antibody that specifically binds to STREP,  
 CC such that the vector delivers the single chain monoclonal antibody coding  
 CC sequence to the cancer cells and the encoded single chain monoclonal  
 CC antibody is expressed intracellularly.  
 CC Note: The present sequence is also shown in sequence listing of the  
 CC specification, but it lacks amino acid residues at its N-terminal end.  
 XX  
 SQ Sequence 375 AA;  
 Query Match 100.0%; Score 116; DB 22; Length 375;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-10;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 REVHPLATSHQOYFYKIPILV 22  
 Db 92 REVHPLATSHQOYFYKIPILV 113  
 RESULT 14  
 AAY11840 ID AAY11840 standard; Protein; 95 AA.  
 XX  
 AC AAY11840;  
 XX  
 DT 18-JUN-1999 (first entry)  
 XX  
 DE Human 5' EST secreted protein SEQ ID No: 440.  
 XX  
 KM Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KM forensic; gene therapy; chromosome mapping; signal peptide; prostate;  
 KM upstream regulatory sequence; cytokine activity; cell proliferation;  
 KM differentiation; haematopoiesis regulation; tissue growth regulation;  
 KM reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KM thrombolytic; anti-inflammatory; tumour inhibition.  
 KM  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9906550-A2.  
 XX  
 PD 11-FEB-1999.

```

PF 31-JUL-1998; 98WO-IB01232.
XX
XX 01-AUG-1997; 97US-0905144.
XX
XX (GENSET ) GENSET.
PA
XX Ductert A, Dumas Milne Edwards J, Lacroix B;
XX PI MPI, 1999-153780/13.
XX DR N-PSDB; AAX40562.
XX
XX New isolated prostate-derived nucleic acids - used to develop
XX products which may have cytokine, immune regulatory, haematopoiesis
XX regulating, anti-inflammatory or tumour inhibition activity
XX
XX Claim 34; Page 577; 675pp; English.
XX
XX AAX40438 to AAX40715 represent 5' expressed sequence tags (ESTs) for
XX human secreted proteins expressed in prostate, and encode the proteins
XX given in AAY11716 to AAY11993 respectively. The proteins given represent
XX the signal peptide and an N-terminal fragment of a secreted protein. The
XX nucleic acid sequences can be used for producing secreted human gene
XX products. They can also be used to develop products for diagnosis and
XX therapy. The proteins obtained may have cytokine activity, cell
XX proliferation and differentiation activity, haematopoiesis regulating
XX activity, tissue growth regulating activity, reproductive hormone
XX regulating activity, chemotactic/chemokinetic activity, haemostatic and
XX thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX activity, tumour inhibition activity or other activities. The products
XX can be used in forensic, gene therapy and chromosome mapping procedures.
XX The sequences can also be used for obtaining corresponding promoter
XX sequences. The nucleic acids encoding the signal peptides can be used for
XX directing extracellular secretion of a polypeptide or the insertion of a
XX polypeptide into a membrane, or importing a polypeptide into a cell.
XX
XX Sequence 95 AA;
SQ
XX
XX Query Match 83.6%; Score 97; DB 20; Length 95;
XX Best Local Similarity 100.0%; Pred. No. 4.2e-08;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 REVYHPLATSHOOYFYKI 18
DB 77 REVYHPLATSHOOYFYKI 94

```

```

PA (GENSET ) GENSET.
XX
XX Ductert A, Dumas Milne Edwards J, Lacroix B;
XX PI MPI, 1999-153778/13.
XX DR N-PSDB; AAX41137.
XX
XX New nucleic acids encoding human secreted proteins - obtained from
XX cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
XX kidney, lung, umbilical cord, placenta and colon tissue
XX
XX Claim 27; Page 677; 824pp; English.
XX
XX AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for
XX human secreted proteins, and encode the proteins given in AAY12261 to
XX AAY12514, respectively. The proteins given represent the signal peptide
XX and an N-terminal fragment of a secreted protein. The nucleic acid
XX sequences can be used for producing secreted human gene products. They
XX can also be used to develop products for diagnosis and therapy. The
XX proteins obtained may have cytokine activity, cell
XX proliferation/differentiation activity, haematopoiesis regulating
XX activity, tissue growth regulating activity, reproductive hormone
XX regulating activity, chemotactic/chemokinetic activity, haemostatic and
XX thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX activity, tumour inhibition activity or other activities. The products
XX can be used in forensic, gene therapy and chromosome mapping procedures.
XX The sequences can also be used for obtaining corresponding promoter
XX sequences. The nucleic acids encoding the signal peptide can be used for
XX directing extracellular secretion of a polypeptide or the insertion of a
XX polypeptide into a membrane, or importing a polypeptide into a cell.
XX
XX Sequence 109 AA;
SQ
XX
XX Query Match 79.3%; Score 92; DB 20; Length 109;
XX Best Local Similarity 94.4%; Pred. No. 3.1e-07;
XX Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 REVYHPLATSHOOYFYKI 18
DB 92 REVYHPLATSHOOYFYKI 109

```

Search completed: March 26, 2003, 16:48:14  
Job time : 56.7209 secs





GenCore version 5.1.4 p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 16:38:14 ; Search time 6.39535 Seconds  
(without alignments)  
142.679 Million cell updates/sec

Title: US-10-010-667A-19

Perfect score: 116  
Sequence: 1 REVHPLATSHQOYFYKIPILV 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.1\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	116	100.0	339	1 STEA_HUMAN	Q9uhes homo sapien
2	52.5	45.3	465	1 C1XG_HAEN	P44458 h citxg pro
3	49	42.2	177	1 MSRA_HAEN	Q9hsg0 halobacteri
4	47.5	40.9	320	1 CYF_PORPU	P51265 porphyra pu
5	46	39.7	474	1 SYFA_ARATH	Q9f034 arabidopsis
6	46	39.7	737	1 YDZT_MYCTU	Q10638 mycobacteri
7	46	39.7	1839	1 CYAA_SACKL	P23466 saccharomyc
8	45	38.8	2549	1 FRAP_MOUSE	Q9j1n9 mus muscula
9	44.5	38.4	137	1 SPMT_PIG	Q28920 sus scrofa
10	44.5	38.4	308	1 PYRB_METAC	Q08122 methanosarc
11	44.5	38.4	1176	1 RP82_DROME	P08266 drosophila
12	44	37.9	811	1 RFXL_YEAST	P48743 saccharomyc
13	44	37.9	928	1 VGLB_HSVBP	P17471 bovine hepr
14	44	37.9	932	1 VGLB_HSVBP	P12640 bovine hepr
15	43.5	37.5	326	1 YJHS_ECOLI	P39370 escherichia
16	43	37.1	337	1 DHAS_VIBCH	P23247 vibrio chol
17	43	37.1	373	1 EXTI_ARATH	Q38913 arabidopsis
18	43	37.1	385	1 GBAS_CABEL	Q20701 caenorhabdi
19	42.5	36.6	299	1 PYRB_ARCFU	Q30130 archaeoglob
20	42	36.2	216	1 MSRA_XYLF	Q9pc45 xyliella fas
21	42	36.2	316	1 DHAS_VIBMI	Q06080 vibrio mimi
22	42	36.2	1437	1 DPO3_BACSV	P13267 bacillus su
23	42	36.2	2549	1 FRAP_HUMAN	P42345 homo sapien
24	42	36.2	2549	1 FRAP_RAT	P42346 rattus norv
25	41.5	35.8	116	1 AONF_PIG	P24020 sus scrofa
26	41.5	35.8	154	1 MYG_ALIMI	P02200 alligator m
27	41	35.3	196	1 COAE_HELPJ	Q92112 helicobacte
28	41	35.3	196	1 COAE_HELPJ	Q25502 helicobacte
29	41	35.3	576	1 SBP_CABEL	Q21950 caenorhabdi
30	41	35.3	623	1 Y014_MYCPN	P75085 mycoplasma
31	40.5	34.9	433	1 XYNB_BOTFI	P26223 butyrivibri
32	40.5	34.9	433	1 PEPP_SALDU	Q85345 salmonella
33	40.5	34.9	492	1 SYN1_YEAST	P25345 saccharomyc

34	40	34.5	115	1	GUAN_RAT	P28902 rattus norv
35	40	34.5	246	1	YNU0_YEAST	P40165 saccharomyc
36	40	34.5	317	1	PLC_ILIMO	P34024 listeria mo
37	40	34.5	394	1	NON9_YEAST	P39107 saccharomyc
38	40	34.5	526	1	C11A_RAT	P14137 rattus norv
39	40	34.5	626	1	PARC_BORBU	O51066 borrelia bu
40	40	34.5	695	1	HRPT_PSESY	P35655 pseudomonas
41	40	34.5	868	1	VGLB_VZVD	P03257 varicella-z
42	40	34.5	870	1	PLSB_XYLF	Q9pe17 xyliella fas
43	40	34.5	903	1	VGLB_HSV1F	P06436 herpes simp
44	40	34.5	904	1	VGLB_HSV1I	P10211 herpes simp
45	40	34.5	904	1	VGLB_HSV1K	P06437 herpes simp

## ALIGNMENTS

```

RESULT 1
ID STEA_HUMAN STANDARD; PRT; 339 AA.
AC Q9UHES; Q95034;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Six transmembrane epithelial antigen of prostate.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20056277; PubMed=10588738;
RA Hubert R.S., Vivanco I., Chen E., Rastegar S., Leong K.,
RA Mitchell S.C., Madraswala R., Zhou Y., Kuo J., Raitano A.B.,
RA Jakobovits A., Salfran D.C., Afari D.E.H.;
RT "STREP: a prostate-specific cell-surface antigen highly expressed in
RT human prostate tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:14523-14526(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Abu-Threideh J., Stoneking T., Langston Y., Maupin R.;
RA Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP TISSUE=Skin;
RC TISSUE=Skin;
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PROSTATE TUMORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF186249; AAF17479.1; -
DR EMBL; AC005053; AAC79150.1; ALT_INT.
DR EMBL; AC004969; AAD15620.1; ALT_INT.
DR EMBL; BC011802; AAH11802.1; -
DR Gene; HGNC:11378; STBP.
KW MIM; 604415; -.
KW Transmembrane; Antigen.
FT TRANSMEM 71 91 POTENTIAL.
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
FT TRANSMEM 218 238 POTENTIAL.
FT TRANSMEM 258 278 POTENTIAL.
FT TRANSMEM 291 311 POTENTIAL.
SQ SEQUENCE 339 AA; 39851 MW; 55443A170C870387 CRC64;

```

Query Match 100.0%; Score 116; DB 1; Length 339;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-11;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REVHPLATSHQOYFYKIPILV 22  
 DB 92 REVHPLATSHQOYFYKIPILV 113

## RESULT 2

CITG\_HAEN STANDARD; PRT; 465 AA.  
 ID CITG\_HAEN  
 AC P44458;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 15-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE CITG protein [includes: Apo-citrate lyase phosphoribosyl-dephospho-CoA transferase (EC 2.7.7.-) (Apo-ACP nucleodityltransferase); 2-(5'-triphosphoribosyl)-3'-dephosphocoenzyme-A synthase (EC 4.2.-.-) (2- (5'-triphosphoribosyl)-3'-dephospho-CoA synthase)].  
 DE CITG OR CITG OR H10021.  
 GN Haemophilus influenzae.  
 OS Haemophilus influenzae; gamma subdivision; Pasteurellaceae;  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 NCBI\_TaxId=727;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Kleibman R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McKenney K., Shetty R., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shteyn R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Utechtack T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Rine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;  
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae RT  
 RL Science 269:496-512 (1995).  
 CC -1- FUNCTION: BI-FUNCTIONAL ENZYME THAT CATALYZES FORMATION OF 2-(5'-TRIPHOSPHORIBOSYL)-3'-DEPHOSPHOCOEZYME-A, AND THEN THE TRANSFER OF THIS PROSTHETIC GROUP PRECURSOR TO THE APO-ACYL CARRIER PROTEIN (GAMMA CHAIN) OF THE CITRATE LYASE TO YIELD THE HOLO-ACYL CARRIER PROTEIN (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: Dephospho-CoA + ATP = 2-(5'-triphosphoribosyl)-3'-dephospho-CoA + adenine.  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE CITG/MDCB FAMILY.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE CITG/MDCB FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: U32687; AAC21699.1; ALT\_INIT.  
 DR TIGR: H10021;  
 DR InterPro: IPR002736; CITG.  
 DR Pfam: PF01874; CitG; 1.  
 KW Transferase; Nucleotidyltransferase; Lyase; Multifunctional enzyme; Complete proteome.  
 FT DOMAIN 1 182 APO-CITRATE LYASE PHOSPHORIBOSYL-DEPHOSPHO-CoA TRANSFERASE.  
 FT DOMAIN 183 465 2-(5'-TRIPHOSPHORIBOSYL)-3'-DEPHOSPHOCOEZYME-A SYNTHASE.  
 FT SEQUENCE 465 AA; 51834 MW; 4467221C3DB58A96 CRC64;

Query Match 45.3%; Score 52.5; DB 1; Length 465;  
 Best Local Similarity 50.0%; Pred. No. 0.66;  
 Matches 10; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 1 REVHPLATSHQOYFYKIP 20  
 DB 83 KEIRPLETGHAEVY-LPI 101

## RESULT 3

MSRA\_HAEN STANDARD; PRT; 177 AA.  
 ID MSRA\_HAEN  
 AC Q9HGO;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Peptide methionine sulfoxide reductase msra (EC 1.8.4.6) (Protein-methionine-S-oxide reductase) (Peptide Met(O) reductase).  
 DE MSRA OR VNG1180G.  
 GN Halobacterium sp. (strain NRC-1).  
 OS Halobacterium sp. (strain NRC-1).  
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 OC Halobacteriaceae; Halobacterium.  
 NCBI\_TaxId=64091;  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M., Shukla H.D., Lasky S.R., Baliga N.S., Thorsen V., Shroya J., Swartzell S., Weir D., Hall J., Dahl T.A., Meli R., Goo Y.A., Leitman B., Keller K., Cruz R., Hanson M.J., Hough D.W., Maddock D.G., Jablonki P.E., Krebs M.P., Angevine C.M., Dale H., Isenbacher T.A., Beck R.P., Pohlischer M., Spudich J.L., Jung K.-H., Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Oner A.D., Braham H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;  
 RA "Genome sequence of Halobacterium species NRC-1.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181 (2000).  
 CC -1- FUNCTION: Has an important function as a repair enzyme for proteins that have been inactivated by oxidation. Catalyzes the reversible oxidation-reduction of methionine sulfoxide in proteins to methionine (by similarity).  
 CC -1- CATALYTIC ACTIVITY: Protein L-methionine + oxidized thioredoxin = protein L-methionine S-oxide + reduced thioredoxin.  
 CC -1- SIMILARITY: BELONGS TO THE MSRA MET SULFOXIDE REDUCTASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: AB005045; AAC19555.1; -.  
 DR HSSP: P54149; IFVA.  
 DR InterPro: IPR002569; PMSR.  
 DR Pfam: PF01625; PMSR; 1.  
 DR ProDom: PD003489; PMSR; 1.  
 DR TIGRFAMs: TIGR00401; msra; 1.  
 KW Oxidoreductase; Complete proteome.  
 FT ACT\_SITE 12 12 BY SIMILARITY.  
 FT SEQUENCE 177 AA; 19202 MW; 337BD2E1CDB8BC CRC64;

Query Match 42.2%; Score 49; DB 1; Length 177;  
 Best Local Similarity 50.0%; Pred. No. 0.81;  
 Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EVHPLATSHQOYFYKIP 19  
 DB 135 ETVFPALEKQNFEEKP 152

## RESULT 4

CYP\_PORPU

```

ID   CYF_PORPU          STANDARD;          PRT;          320 AA.
AC   P51265;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Apocytochrome F precursor.
GN   PETA.
OS   Porphyra purpurea.
OC   Chloroplast.
OC   Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX   NCBI_TaxId=2787;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Avonport;
RA   Reich M.E., Munholland J.;
RT   "Complete nucleotide sequence of the Porphyra purpurea chloroplast
    genome."
RL   Plant Mol. Biol. Rep. 13:333-335(1995).
CC   -1- FUNCTION: TRANSLOCATES PROTONS ACROSS THE THYLAKOID MEMBRANE AND
CC   TRANSFERS ELECTRONS FROM PHOTOSYSTEM II TO PHOTOSYSTEM I. IT
CC   RECEIVES ELECTRONS FROM THE RISKKE IRON-SULFUR PROTEIN AND PASSES
CC   THEM TO PLASTOCYANIN. THIS FUNCTION IS VERY SIMILAR TO THAT OF
CC   MITOCHONDRIAL CYTOCHROME C1.
CC   -1- SUBUNIT: MEMBER OF THE CYTOCHROME B6/F COMPLEX INCLUDING
CC   CYTOCHROME B6, CYTOCHROME F AND PROBABLY AN IRON SULFUR PROTEIN.
CC   -1- SIMILARITY: BELONGS TO THE CYTOCHROME C FAMILY.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; U38804; AAC08151.1; -.
DR   HSSP; P36438; 1HCZ.
DR   InterPro; IPR002325; Apocyt F.
DR   InterPro; IPR000345; Cyt_c_heme_bind.
DR   Pfam; PF01333; Apocytocytome_F; 1.
DR   PRINTS; PR00610; CYTOCHROME_F.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Electron transport; Heme; Chloroplast; Thylakoid;
KW   Photosynthesis; Photosystem I; Photosystem II; Transit peptide;
KW   Transmembrane.
FT   TRANSIT 1..36  CHLOROPLAST (BY SIMILARITY).
FT   CHAIN 37..320  APOCYTOCHROME F.
FT   BINDING 57..57  HEME (COVALENT) (PROBABLE).
FT   BINDING 60..60  HEME (COVALENT) (PROBABLE).
FT   METAL 61..61  IRON (HEME AXIAL LIGAND) (PROBABLE).
FT   TRANSMEM 286..305  POTENTIAL.
SQ   SEQUENCE 320 AA; 35012 MW; C9B103F0453369C7 CRC64;
Query Match 40.9%; Score 47.5; DB 1; Length 320;
Best Local Similarity 40.7%; Pred. No. 2.8;
Matches 11; Conservative 4; Mismatches 7; Indels 5; Gaps 1;
Qy 1 REVIMPL-----ATSHQOYFKIPILV 22
Db 160 REIIFILSPDPKAKQAHFKFYPIYV 186

```

```

OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsie.
OX   NCBI_TaxId=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083488; PubMed=10617198;
RA   Mayer K.F.X., Scheller C., Wambutt R., Murphy G., Volckaert G.,
RA   Pohl T., Duesterhoeft A., Striekema M., Ertlan K.-D., Teyrn N.,
RA   Harris B., Amorge W., Brandt P., Grivell L.A., Rieger M., Mueller M.,
RA   Weichselgartner M., de Simone V., Oberwaller B., Macho R., Schmidtkei T.,
RA   Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA   Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA   Vos P., Hohensei J., Zimmermann W., Medler H., Ridley P.,
RA   Langham S.-A., McCullagh B., Biham L., Robben J.,
RA   Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandebussche F.,
RA   Braeken M., Weltjens I., Voet M., Baetleins I., Aert R., Delfor E.,
RA   Weltzenegger T., Bothe G., Rameberger U., Hilbert H., Braun M.,
RA   Holzner B., Brandt A., Peters S., van Staveren M., Dirxse W.,
RA   Moollman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA   Bernerster S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA   De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
RA   Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA   Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,
RA   Pettett A., Rajandream M.A., Lyne M., Benes V., Reckmann S.,
RA   Borkova D., Bioecker H., Scharfe M., Grimm M., Loehner T.-H.,
RA   Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA   Gabel G., Fuchs M., Fartman B., Grandeneth K., Danner D., Hetzl A.,
RA   Neumann S., Argitiro A., Vitale D., Lignotti R., Piravandi E.,
RA   Massenet O., Quigley F., Clabaud G., Muendlein A., Felber R.,
RA   Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA   Cheffor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA   Gibbons T., Weber N., Vandenbol M., Bargues M., Terol J., Torres A.,
RA   Perez-Perez A., Fumelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA   Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bietke C.,
RA   Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA   Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA   Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA   Sektion M., Murray J., Shee P., Cordes M., Abu-Threiden J.,
RA   Stenking T., Kallick J., Graves T., Harmon G., Edwards J.,
RA   Latraille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA   Min P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA   Nelson J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA   Nelson J., Splich J., Ryan E., Andrews S., Geisel C., Layman D.,
RA   Du H., Ali J., Berghoff A., Jones K., Dione K., Cotton M., Joshi C.,
RA   Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA   Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,
RA   Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA   Grant S., Shobdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA   Chen B., Maria M., Martienssen R., McCombie W.R.;
RT   "Sequence and analysis of chromosome 4 of the plant Arabidopsis
    thaliana."
RL   Nature 402:769-777(1999).
CC   -1- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
CC   dihydrophosphate + L-phenylalanyl-tRNA(Phe).
CC   -1- SUBUNIT: Tetramer of two alpha and two beta chains (by
CC   similarity).
CC   -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC   -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC   PHE-TRNA SYNTHETASE ALPHA CHAIN SUBFAMILY 2.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; AL050351; CAB43643.1; ALT_SEQ.
DR   EMBL; AL051594; CAB80591.1; ALT_SEQ.

```

DR HSSP; P27001; 1PVS.  
 DR InterPro; IPR002106; AACRNA\_ligaseII.  
 DR InterPro; IPR004529; PheS.  
 DR InterPro; IPR002319; tRNA-synt\_2d.  
 DR Pfam; PF01409; tRNA-synt\_2d; 1.  
 DR tIGRFam; tIGR00468; pheS; 1.  
 DR PROSITE; PS00862; AA\_tRNA\_LIGASE\_II; 1.  
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.  
 SQ SEQUENCE 474 AA; 54516 MW; 0BF7F1BA07C03C77 CRC64;  
 Query Match 39.7%; Score 46; DB 1; Length 474;  
 Best Local Similarity 46.7%; Pred. No. 7.5;  
 Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;  
 Oy 5 HPLATSHOQFYKIP 19  
 Db 246 HPARSDHDTFFLKIP 260  
 RESULT 6  
 ID YD27 MYCTU STANDARD; PRT; 737 AA.  
 AC Q10638;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein RV1327C.  
 GN RV1327C OR MT1369 OR MTCV130.12C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 NCBI\_TaxID=1773;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=98295987; Pubmed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sullivan J.E., Taylor K., Whitehead S., Barrett B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z73902; CAA98091.1; -.  
 DR EMBL; AEG07010; AAK45633.1; -.  
 DR TIGR; MT1369; -.  
 DR TubercuList; RV1327C; -.  
 DR InterPro; IPR000461; Alpha\_amylase.  
 DR Pfam; PF00128; alpha-amylase; 1.

KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 737 AA; 82805 MW; 567EE418B391F1F CRC64;  
 Query Match 39.7%; Score 46; DB 1; Length 737;  
 Best Local Similarity 46.7%; Pred. No. 12;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 Oy 5 HPLATSHOQFYKIP 19  
 Db 389 HPWABRHQWTFELP 403  
 RESULT 7  
 ID CYAA SACKL STANDARD; PRT; 1839 AA.  
 AC P23456;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Adenylate cyclase (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenylyl  
 DE cyclase).  
 GN CYRL.  
 OS Saccharomyces kluyveri (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 NCBI\_TaxID=4934;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91323718; Pubmed=1864503;  
 RA Young D., O'Neill K., Broek D., Wiggler M.;  
 RT "The adenyllyl cyclase-encoding gene from Saccharomyces kluyveri.";  
 RL Gene 102:129-132(1991).  
 CC -1- FUNCTION: PLAYS ESSENTIAL ROLES IN REGULATION OF CELLULAR  
 CC METABOLISM BY CATALYSING THE SYNTHESIS OF A SECOND MESSENGER,  
 CC CAMP.  
 CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.  
 CC -1- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-3 FAMILY.  
 CC -1- SIMILARITY: CONTAINS 21 LEUCINE-RICH REPEATS (LRR).  
 CC -1- SIMILARITY: CONTAINS 1 PP2C-LIKE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X56042; CAA39513.1; -.  
 DR PIR; S14464; OYBYK.  
 DR PIR; J01145; J01145.  
 DR InterPro; IPR001054; G\_cyclase.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR003592; LRR\_out.  
 DR InterPro; IPR003591; LRR\_cyp.  
 DR InterPro; IPR001932; PP2C-like.  
 DR InterPro; IPR000159; RA domain.  
 DR Pfam; PF00211; Guanylate\_cyc; 1.  
 DR Pfam; PF00481; PP2C; 1.  
 DR Pfam; PF00560; LRR; 13.  
 DR PRINTS; PR00019; LEURICHRPT.  
 DR SMART; SM00044; CYCC; 1.  
 DR SMART; SM00370; LRR; 4.  
 DR SMART; SM00369; LRR\_TYP; 3.  
 DR SMART; SM00332; PP2Cc; 1.  
 DR SMART; SM00314; RA; 1.  
 DR PROSITE; PS0125; GUANYLATE\_CYCLASES\_2; 1.  
 KW Lyase; Repeat; Leucine-rich repeat; CAMP synthesis; Magnesium.  
 DR REPEAT 632 655 LRR 1.  
 FT REPEAT 632 655 LRR 2.  
 FT REPEAT 659 679 LRR 3.  
 FT REPEAT 680 702 LRR 4.  
 FT REPEAT 703 724 LRR 5.  
 FT REPEAT 726 748 LRR 5.

```

FT REPEAT 749 771 LRR 6.
FT REPEAT 773 793 LRR 7.
FT REPEAT 794 818 LRR 8.
FT REPEAT 820 833 LRR 9.
FT REPEAT 834 857 LRR 10.
FT REPEAT 859 879 LRR 11.
FT REPEAT 880 903 LRR 12.
FT REPEAT 905 926 LRR 13.
FT REPEAT 927 949 LRR 14.
FT REPEAT 951 974 LRR 15.
FT REPEAT 1004 1025 LRR 16.
FT REPEAT 1026 1048 LRR 17.
FT REPEAT 1050 1072 LRR 18.
FT REPEAT 1074 1096 LRR 19.
FT REPEAT 1101 1124 LRR 20.
FT REPEAT 1135 1160 LRR 21.
FT DOMAIN 1185 1440 PRPC-LIKE.
FT DOMAIN 1441 1839 CATALYTIC.
SQ SEQUENCE 1839 AA; 206695 MW; 86A69BCB1F273CB CRC64;

```

```

Query Match 39.7%; Score 46; DB 1; Length 1839;
Best Local Similarity 66.7%; Pred. No. 35;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 3 VHPPLATSHOOY 14
DB 589 VHPPLATSHOY 600

```

```

RESULT 8
FRAP_MOUSE STANDARD; PRT; 2549 AA.
ID FRAP_MOUSE
AC 09JUN9;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE PKBP-rapamycin associated protein (FRAP).
GN FRAP1 OR FRAP.
OS Mus musculus (Mouse).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxId=10090;
[1]
SEQUENCE FROM N.A.

```

```

RA Bliskovsky V., Mock B.;
RT "Positional cloning of mouse plasmacytoma susceptibility gene.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ACTS AS THE TARGET FOR THE CELL-CYCLE ARREST AND
CC IMMUNOSUPPRESSIVE EFFECTS OF THE PKBP12-RAPAMYCIN COMPLEX.
CC -!- SIMILARITY: BELONGS TO THE PI3/P14-KINASES FAMILY.
CC -!- SIMILARITY: CONTAINS 8 HEAT REPEATS.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; AF152838; AAF73196.1; -
DR HSSP; PA2345; 1FAP.
DR MGD; MGI:1928394; FRAP1.
DR InterPro; IPR003151; FAT.
DR InterPro; IPR003152; FATC.
DR InterPro; IPR000357; HEAT_repeat.
DR InterPro; IPR000403; PI3_P14_kinase.
DR Pfam; PF00454; PI3_P14_kinase; 1.
DR Pfam; PF02259; FAT_1.
DR Pfam; PF02260; FATC_1.
DR SMART; SM00146; PI3Kc_1.
DR PROSITE; PS00915; PI3_4_KINASE_1; 1.

```

```

DR PROSITE; PS00916; PI3_4_KINASE_2; 1.
DR PROSITE; PS00290; PI3_4_KINASE_3; 1.
DR PROSITE; PS0077; HEAT_REPEAT; FALSE_NEG.
KW Transferase; Kinase; Repeat.
FT REPEAT 16 53 HEAT 1.
FT REPEAT 650 688 HEAT 2.
FT REPEAT 859 897 HEAT 3.
FT REPEAT 988 1025 HEAT 4.
FT REPEAT 1069 1106 HEAT 5.
FT REPEAT 1109 1148 HEAT 6.
FT REPEAT 1150 1186 HEAT 7.
FT DOMAIN 1382 1982 FAT.
FT DOMAIN 1933 1970 HEAT 8.
FT DOMAIN 2182 2549 PI3K/PI4K.
SQ SEQUENCE 2549 AA; 288734 MW; C826B35E861BC3 CRC64;

```

```

Query Match 38.8%; Score 45; DB 1; Length 2549;
Best Local Similarity 50.0%; Pred. No. 73;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 REVHPPLATSHOOYFY 16
DB 1683 RQDHPPLATSHOYFY 1698

```

```

RESULT 9
SPMI_PIG STANDARD; PRT; 137 AA.
ID SPMI_PIG
AC Q28920;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Seminal plasma sperm motility inhibitor precursor.
GN SPMI.
OS Sus scrofa (Pig).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
NCBI_TaxId=9823;
[1]
SEQUENCE FROM N.A.

```

```

RC TISSUE=Seminal vesicle;
RX MEDLINE=95361914; PubMed=7635190;
RA Iwanoto T., Hiroaki H., Furuichi Y., Wada K., Satoh M., Satoh M.,
RA Osada T., Gagnon C.;
RT "Cloning of boar SPMI gene which is expressed specifically in seminal
RT vesicle and codes for a sperm motility inhibitor protein.";
RL FEBS Lett. 368:420-424(1995).
CC -!- FUNCTION: INHIBITOR OF SPERM MOTILITY.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: SEMINAL PLASMA OR SPERM.
CC -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE SPERMADHESIN FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; S80568; AAB35000.2; -
DR HSSP; P29392; 1SPF.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000124; Spermadhesin.
DR Pfam; PF00431; CUB; 1.
DR SMART; SM00042; CUB; 1.
DR PROSITE; PS00985; SPERMADHESIN_1; 1.
DR PROSITE; PS00986; SPERMADHESIN_2; FALSE_NEG.
DR PROSITE; PS01180; CUB; 1.
DR Heparin-binding; Glycoprotein; Fertilization; Signal.
KW Heparin-binding; Glycoprotein; Fertilization; Signal.
FT SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 137 SEMINAL PLASMA SPERM MOTILITY INHIBITOR.

```

FT DOMAIN 30 131 CUB.  
 FT DISULFID 30 51 BY SIMILARITY.  
 FT CARBOHYD 36 36 N-LINKED (GLCNAC... ) (POTENTIAL).  
 SO SEQUENCE 137 AA; 15194 MM; 9DC107799AAC71EF CRC64;  
 Query Match 38.4%; Score 44.5; DB 1; Length 137;  
 Best Local Similarity 57.9%; Pred. No. 3.2;  
 Matches 11; Conservative 1; Mismatches 6; Indels 1; Gaps 1;  
 QY 1 REV1HPLATSHOQFYKIP 19  
 DB 115 RESHHP-ASSFNVFYGP 132  
 RESULT 10  
 PYRB\_METAC STANDARD; PRT; 308 AA.  
 ID PYRB\_METAC  
 AC Q8THL2;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Aspartate carbamoyltransferase (EC 2.1.3.2) (Aspartate  
 transcarbamylase) (ATCase).  
 GN PYRB OR M44502.  
 OS Methanosarcina acetivorans.  
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
 OC Methanosarcinaceae; Methanosarcina.  
 NC NCB1\_TaxID=2214;  
 RX MEDLINE=21929760; PubMed=11932238;  
 RA Galagan J.E., Nusbbaum C., Roy A., Smirnov S.G., Macdonald P.,  
 Fitzhugh W., Calvo S., Engels R., Smitrov M.G., Altmann D., Brown A.,  
 Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,  
 Limon L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,  
 Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,  
 Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,  
 Leigh J.A., Li W., Liu J., Mckhopadhyay B., Reeve J.N., Smith K.,  
 Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,  
 Ferry J.G., Jarrell K.F., Jjing H., Macario A.J.L., Paulsen I.,  
 Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,  
 Metcalf W.W., Birren B.  
 RA "The genome of Methanosarcina acetivorans reveals extensive metabolic  
 and physiological diversity";  
 RT Genome Res. 12:532-543(2002).  
 RL  
 CC -1- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-aspartate = phosphate  
 + N-carbamoyl-L-aspartate.  
 CC -1- PATHWAY: Pyrimidine biosynthesis; second step.  
 CC -1- SUBUNIT: Heterooligomer of catalytic and regulatory chains.  
 CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AEO11170; AAM07842.1;  
 DR PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.  
 DR Pyrimidine biosynthesis; Transferase; Complete proteome.  
 SO SEQUENCE 308 AA; 34644 MM; 33282B48E28595C CRC64;  
 Query Match 38.4%; Score 44.5; DB 1; Length 308;  
 Best Local Similarity 35.3%; Pred. No. 8.1;  
 Matches 12; Conservative 4; Mismatches 3; Indels 15; Gaps 2;  
 QY 2 BV1HPL-----ATSH-----QQFYKIP 20  
 DB 262 KV1HPLRVNEIADVDATPACVFEQAFYGVPI 295

RESULT 11  
 RPB2\_DROME STANDARD; PRT; 1176 AA.  
 ID RPB2\_DROME  
 AC P08266; Q95027; Q04155; Q9VFM7;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA-directed RNA polymerase II 140 kDa polypeptide (EC 2.7.7.6)  
 DE (RNA polymerase II subunit 2).  
 GN RPI140 OR CG3180.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 NC NCB1\_TaxID=7227;  
 RX MEDLINE=88011299; PubMed=3116266;  
 RA Falkenburg D., Dworniczak B., Faust D.M., Bautz E.K.F.;  
 RT "RNA polymerase II of Drosophila. Relation of its 140,000 Mr subunit  
 to the beta subunit of Escherichia coli RNA polymerase";  
 RL J. Mol. Biol. 195:929-937(1987).  
 RN (2)  
 RP SEQUENCE OF 1-69 FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=91276237; PubMed=1905256;  
 RA Stitzler S., Oldenburg I., Peterson G., Bautz E.K.F.;  
 RT "Analysis of the promoter region of the housekeeping gene DmRP140 by  
 RT sequence comparison of Drosophila melanogaster and Drosophila  
 RT virilis";  
 RL Gene 100:155-162(1991).  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkely;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 Amandes P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 Wan K.H., Doyle E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 Abril J.F., Agbayani A., An H.-J., Andrews-Planck D., Baldwin D.,  
 Ballieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 Borokova D., Botchan M.R., Butler J., Brokstein P., Brotier P.,  
 Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 DePamphilis M., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 Fosler A., Gabrielian A.E., Gang N.S., Gelbart W.M., Glasser K.,  
 Glosk A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeigbam C.,  
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,  
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 Laeko P., Lai Y., Levyteky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostreft A.,  
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 Nelson D.R., Nelson K.A., Nixon K.A., Nusskern D.R., Pacleb J.M.,  
 Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 Stryker E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Weissman T.A., Weissman G.M., Weissbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;



```

FT DOMAIN 68 780 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 781 801 POTENTIAL.
FT CARBOHYD 804 824 POTENTIAL.
FT DOMAIN 825 928 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 442 442 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 484 484 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 579 579 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 637 637 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 703 703 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 409 409 S -> T (IN REF. 1; AAA46013).
FT CONFLICT 673 673 G -> P (IN REF. 1; AAA46013).
SQ SEQUENCE 928 AA; 102177 MW; B47982224PCD769D CRC64;

Query Match
Best Local Similarity 37.9%; Score 44; DB 1; Length 928;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 REVHPLATSHOQYF 15
Db 656 RELVEPCTANHKRYF 670

RESULT 14
VGLB HSVBC STANDARD; PRT; 932 AA.
ID VGLB HSVBC
AC P12640;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glycoprotein I precursor (Glycoprotein GVP-6) (glycoprotein IIA)
DE (Glycoprotein 16) (Glycoprotein G130) (Glycoprotein B).
GN GI OR U127.
OS Bovine herpesvirus type 1 (strain Cooper).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10323;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8830884; PubMed=2841484;
RA Whitebeck J.C., Bello U.J., Lawrence W.C.;
RT "Comparison of the bovine herpesvirus 1 g1 gene and the herpes
" complex virus type 1 gB gene."
RL J. Virol. 62:3319-3327(1988).
CC -1- SUBUNIT: DIMER. PROBABLY LINKED BY DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC -----
DR EMBL; M21474; AAA46055.1; -
DR EMBL; Z78205; CAB01598.1; -
DR EMBL; AJ004601; CA06106.1; -
DR PIR; A28877; VGBBEC.
DR InterPro; IPR000234; Glycoprot_B.
DR Pfam; PF00606; Glycoprotein_B_1.
DR Prodom; PD000693; Glycoprot_B_1.
KW Glycoprotein; Transmembrane; Signal.
FT CHAIN 1 67
FT SIGNAL 68 932 GLYCOPROTEIN I.
FT DOMAIN 68 758 EXTRACELLULAR.
FT TRANSMEM 759 827 POTENTIAL.
FT DOMAIN 828 932 CYTOPLASMIC.
FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 640 640 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 706 706 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 932 AA; 101195 MW; 9DCEA85C5FC3DA3 CRC64;

Query Match
Best Local Similarity 37.9%; Score 44; DB 1; Length 932;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 REVHPLATSHOQYF 15
Db 659 RELVEPCTANHKRYF 673

RESULT 15
YJHS ECOLI STANDARD; PRT; 326 AA.
ID YJHS ECOLI
AC P39370;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yjhs precursor.
DE YJHS OR B4309.
GN Escherichia coli.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burdett V.D., Plunkett G. III, Sofia H.U., Daniels D.L.,
RA Blatner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes."
RL Nucleic Acids Res. 23:2105-2119(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC -----
DR EMBL; U14003; AAA97205.1; -
DR EMBL; AE000501; AAC77265.1; -
DR BioGene; BG12561; YJHS.
DR InterPro; IPR005181; DUF303.
DR Pfam; PF03629; DUF303; 1.
KW Hypothetical protein; Signal; Complete proteome.
FT CHAIN 1 21
FT SIGNAL 22 326 HYPOTHETICAL PROTEIN YJHS.
SQ SEQUENCE 326 AA; 36878 MW; 6E48217701678DAD CRC64;

Query Match
Best Local Similarity 37.5%; Score 43.5; DB 1; Length 326;
Matches 9; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 5 HPLATSHO-QY 14
Db 80 HPLATSHOQY 90

Search completed: March 26, 2003, 16:45:05
Job time : 8.39535 secs

```



GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: March 26, 2003, 16:49:57 ; Search time 14.8372 Seconds  
(without alignments)  
87.073 Million cell updates/sec

Title: US-10-010-667A-19  
Perfect score: 116  
Sequence: 1 REVHPLATSHQOYFYKIPILV 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 237916 seqs, 58723674 residues

Total number of hits satisfying chosen parameters: 237916

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubppaa/PCR\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubppaa/PCRUS\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubppaa/US05\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	116	100.0	22	9	US-10-011-095-19
2	116	100.0	22	9	US-10-010-667A-19
3	116	100.0	104	10	US-09-864-761-47234
4	116	100.0	267	10	US-09-747-835A-50
5	116	100.0	267	10	US-09-747-835A-51
6	116	100.0	339	9	US-10-012-896-879
7	116	100.0	339	9	US-09-802-520-11
8	116	100.0	339	9	US-09-895-793-879
9	116	100.0	339	9	US-09-895-814-879
10	116	100.0	339	9	US-10-011-095-2
11	116	100.0	339	9	US-10-010-667A-2
12	116	100.0	339	10	US-09-759-143-879
13	116	100.0	339	10	US-09-780-669-879
14	116	100.0	339	10	US-09-822-827-879
15	68	58.6	490	9	US-09-802-520-1
16	48	41.4	237	10	US-09-747-835A-15
17	48	41.4	488	10	US-09-747-835A-13
18	42	36.2	314	10	US-09-815-242-10794
19	42	36.2	1140	9	US-09-950-634-4

20	42	36.2	2549	9	US-09-950-634-3	Sequence 3, Appli
21	41	35.3	469	10	US-09-866-582-10	Sequence 10, Appl
22	40	34.5	59	9	US-10-097-065-147	Sequence 147, App
23	40	34.5	206	12	US-10-127-733-2	Sequence 2, Appli
24	40	34.5	359	9	US-10-055-364-48	Sequence 48, Appl
25	40	34.5	430	10	US-09-799-777-62	Sequence 62, Appl
26	40	34.5	437	9	US-09-738-626-4226	Sequence 4226, Ap
27	40	34.5	437	10	US-09-746-660A-2	Sequence 2, Appli
28	40	34.5	437	10	US-09-919-937-32	Sequence 344, App
29	40	34.5	1536	9	US-10-043-487-344	Sequence 20, Appl
30	39.5	34.1	133	10	US-09-734-569-20	Sequence 45491, A
31	39	33.6	64	10	US-09-864-761-45491	Sequence 1, Appli
32	39	33.6	339	12	US-10-067-291-1	Sequence 5590, Ap
33	39	33.6	636	9	US-10-007-343-2	Sequence 32, Appl
34	39	33.6	553	9	US-09-738-626-5590	Sequence 16, Appl
35	39	33.6	777	9	US-09-945-901-16	Sequence 16, Appl
36	39	33.6	777	9	US-10-007-747-16	Sequence 16, Appl
37	39	33.6	777	9	US-10-038-937-16	Sequence 32, Appl
38	39	33.6	854	9	US-09-945-901-32	Sequence 32, Appl
39	39	33.6	854	9	US-10-007-747-32	Sequence 30, Appl
40	39	33.6	854	9	US-10-038-937-32	Sequence 30, Appl
41	39	33.6	870	9	US-09-945-901-30	Sequence 30, Appl
42	39	33.6	870	9	US-10-007-747-30	Sequence 30, Appl
43	39	33.6	870	9	US-10-038-937-30	Sequence 30, Appl
44	39	33.6	875	9	US-09-945-901-28	Sequence 28, Appl
45	39	33.6	875	9	US-10-007-747-28	Sequence 28, Appl

## ALIGNMENTS

RESULT 1  
US-10-011-095-19  
; Sequence 19, Application US/10011095  
; Publication NO. US20030045682A1  
; GENERAL INFORMATION:  
; APPLICANT: Afar, Daniel  
; APPLICANT: Hubert, Rene S.  
; APPLICANT: Leong, Kahan  
; APPLICANT: Raitano, Arthur B.  
; APPLICANT: Saffran, Douglas C.  
; APPLICANT: Mitchell, Steve Chappell  
; TITLE OF INVENTION: ANTIBODIES IMMUNOSPECIFIC FOR STEAP1 (AS AMENDED)  
; FILE REFERENCE: 511582001610  
; CURRENT APPLICATION NUMBER: US/10/011,095  
; CURRENT FILING DATE: 2001-12-06  
; PRIOR APPLICATION NUMBER: 09/323,873  
; PRIOR FILING DATE: 1999-06-01  
; PRIOR APPLICATION NUMBER: 60/087,520  
; PRIOR FILING DATE: 1998-06-01  
; PRIOR APPLICATION NUMBER: 60/091,183  
; PRIOR FILING DATE: 1998-06-30  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 19  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: STEAP-1 peptide  
; US-10-011-095-19

Query Match 100.0%; Score 116; DB 9; Length 22;  
Best Local Similarity 100.0%; Pred. No. 7.4e-12;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REVHPLATSHQOYFYKIPILV 22  
|||  
Db 1 REVHPLATSHQOYFYKIPILV 22

RESULT 2  
US-10-010-667A-19

```
/ Sequence 19, Application US/10010667A
/ Publication No. US20030055217A1
/ GENERAL INFORMATION:
/ APPLICANT: Afar, Daniel
/ APPLICANT: Hubert, Rene S.
/ APPLICANT: Leong, Kahen
/ APPLICANT: Raitano, Arthur B.
/ APPLICANT: Saffran, Douglas C.
/ APPLICANT: Mitchell, Steve Chappell
/ TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
/ TITLE OF INVENTION: EXPRESSED IN HUMAN CANCERS AND USES THEREOF
/ FILE REFERENCE: 511582001601
/ CURRENT APPLICATION NUMBER: US/10/010,667A
/ PRIOR FILING DATE: 2001-12-06
/ PRIOR APPLICATION NUMBER: 09/323,873
/ PRIOR FILING DATE: 1999-06-01
/ PRIOR APPLICATION NUMBER: 60/087,520
/ PRIOR FILING DATE: 1998-06-01
/ PRIOR APPLICATION NUMBER: 60/091,183
/ PRIOR FILING DATE: 1998-06-30
/ NUMBER OF SEQ ID NOS: 32
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 19
/ LENGTH: 22
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: STEAP-1 peptide
/ US-10-010-667a-19
```

```
Query Match 100.0%; Score 116; DB 9; Length 22;
Best Local Similarity 100.0%; Pred. No. 7,4e-12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 REVYHPLATSHQGYFYKIPILV 22
Db 1 REVYHPLATSHQGYFYKIPILV 22
```

```
RESULT 3
US-09-864-761-47234
/ Sequence 47234, Application US/09864761
/ Patent No. US20020048763A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ APPLICANT: Chen, Wensheng
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/ TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
/ FILE REFERENCE: Aeomica-X-1
/ CURRENT APPLICATION NUMBER: US/09/864,761
/ CURRENT FILING DATE: 2001-05-23
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2000-02-04
/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263,6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
```

```
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 47234
/ LENGTH: 104
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AC005053.1
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.87
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.57
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.59
/ OTHER INFORMATION: SWISSPROT HIT: O67305, EVALU 5.60e+00
/ OTHER INFORMATION: EST_HUMAN HIT: BE875216.1, EVALU 2.00e-54
/ US-09-864-761-47234
```

```
Query Match 100.0%; Score 116; DB 10; Length 104;
Best Local Similarity 100.0%; Pred. No. 4,4e-11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 REVYHPLATSHQGYFYKIPILV 22
Db 59 REVYHPLATSHQGYFYKIPILV 80
```

```
RESULT 4
US-09-747-835A-50
/ Sequence 50, Application US/09747835A
/ Patent No. US2002014692A1
/ GENERAL INFORMATION:
/ APPLICANT: Yamazaki, Victoria
/ APPLICANT: Tang, Y. Tom
/ APPLICANT: Liu, Chenghua
/ APPLICANT: Zhou, Ping
/ APPLICANT: Wang, Dunrui
/ APPLICANT: Zhang, Jie
/ APPLICANT: Ren, Feiyan
/ APPLICANT: Asundi, Vinod
/ APPLICANT: Dimauc, Radoje T
/ TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-LIKE
/ TITLE OF INVENTION: LIKE POLYPEPTIDES AND POLYNUCLEOTIDES
/ FILE REFERENCE: HVS-37CIP
/ CURRENT APPLICATION NUMBER: US/09/747,835A
/ CURRENT FILING DATE: 2002-03-08
/ PRIOR APPLICATION NUMBER: US 09/729,739
/ PRIOR FILING DATE: 2000-12-04
/ PRIOR APPLICATION NUMBER: US 09/653,450
/ PRIOR FILING DATE: 2000-08-31
/ PRIOR APPLICATION NUMBER: US 09/620,312
/ PRIOR FILING DATE: 2000-07-19
/ PRIOR APPLICATION NUMBER: US 09/598,042
/ PRIOR FILING DATE: 2000-06-20
/ PRIOR APPLICATION NUMBER: US 09/552,317
/ PRIOR FILING DATE: 2000-04-25
/ PRIOR APPLICATION NUMBER: US 09/488,725
/ PRIOR FILING DATE: 2000-01-21
/ NUMBER OF SEQ ID NOS: 63
```

SOFTWARE: Patentin version 3.0  
SEQ ID NO 50  
LENGTH: 267  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-747-835A-50

Query Match 100.0%; Score 116; DB 10; Length 267;  
Best Local Similarity 100.0%; Pred. No. 1.3e-10;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 REVHPLATSHOOYFYKIPILV 22  
Db 26 REVHPLATSHOOYFYKIPILV 47

RESULT 5  
US-09-747-835A-51  
Sequence 51, Application US/09747835A  
Patent No. US2002014692A1  
GENERAL INFORMATION:  
APPLICANT: Yamazaki, Victoria  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Zhou, Ping  
APPLICANT: Wang, Dunru  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Asundi, Vinod  
APPLICANT: Drmanac, Radoje T  
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-LIKE  
FILE REFERENCE: HYS-37CIP  
CURRENT APPLICATION NUMBER: US/09/747, 835A  
PRIOR FILING DATE: 2002-03-08  
PRIOR APPLICATION NUMBER: US 09/729, 739  
PRIOR FILING DATE: 2000-12-04  
PRIOR APPLICATION NUMBER: US 09/653, 450  
PRIOR FILING DATE: 2000-08-31  
PRIOR APPLICATION NUMBER: US 09/620, 312  
PRIOR FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: US 09/598, 042  
PRIOR FILING DATE: 2000-06-20  
PRIOR APPLICATION NUMBER: US 09/552, 317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: US 09/488, 725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 51  
LENGTH: 267  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-747-835A-51

Query Match 100.0%; Score 116; DB 10; Length 267;  
Best Local Similarity 100.0%; Pred. No. 1.3e-10;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 REVHPLATSHOOYFYKIPILV 22  
Db 26 REVHPLATSHOOYFYKIPILV 47

RESULT 6  
US-10-012-896-879  
Sequence 879, Application US/10012896  
Publication No. US20020183251A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Jianshun  
APPLICANT: Dillon, Davin C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Harlocker, Susan L.

APPLICANT: Jiang, Yugu  
APPLICANT: Kalos, Michael D.  
APPLICANT: Retter, Marc W.  
APPLICANT: Stolk, John A.  
APPLICANT: Day, Craig H.  
APPLICANT: Vedwick, Thomas S.  
APPLICANT: Carter, Darrick  
APPLICANT: Li, Samuel X.  
APPLICANT: Wang, Ajun  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Hepler, William T.  
APPLICANT: Henderson, Robert A.  
APPLICANT: Hural, John  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Houghton, Raymond L.  
APPLICANT: Vinals de Bassols, Carlota  
APPLICANT: Foy, Teresa  
APPLICANT: Ranger, Gary R.  
APPLICANT: Wantanabe, Yoshinhiro  
APPLICANT: Meagher, Madeleine Joy  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
FILE REFERENCE: 210121.427C27  
CURRENT APPLICATION NUMBER: US/10/012, 896  
PRIOR FILING DATE: 2001-12-10  
NUMBER OF SEQ ID NOS: 1011  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 879  
LENGTH: 339  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-012-896-879

Query Match 100.0%; Score 116; DB 9; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1.7e-10;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 REVHPLATSHOOYFYKIPILV 22  
Db 92 REVHPLATSHOOYFYKIPILV 113

RESULT 7  
US-09-802-520-11  
Sequence 11, Application US/09802520  
Publication No. US20020187472A1  
GENERAL INFORMATION:  
APPLICANT: Lal, Preeti  
APPLICANT: Faris, Mary  
APPLICANT: Chen, Hui-Mei  
APPLICANT: Ison, Craig H.  
TITLE OF INVENTION: STREP-RELATED PROTEIN  
FILE REFERENCE: PC-0037 US  
CURRENT APPLICATION NUMBER: US/09/802, 520  
PRIOR FILING DATE: 2001-03-09  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PERL Program  
SEQ ID NO 11  
LENGTH: 339  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. US20020187472A1 96572948  
US-09-802-520-11

Query Match 100.0%; Score 116; DB 9; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1.7e-10;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 REVHPLATSHOOYFYKIPILV 22  
Db 92 REVHPLATSHOOYFYKIPILV 113

```
RESULT 8
US-09-895-793-879
; Sequence 879, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Derrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Horai, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Ranger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895,793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 879
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-793-879

Query Match 100.0%; Score 116; DB 9; Length 339;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REVHPLATSHOOYFYKIPILV 22
Db 92 REVHPLATSHOOYFYKIPILV 113

RESULT 9
US-09-895-814-879
; Sequence 879, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Derrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
```

```
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Ranger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 879
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-814-879

Query Match 100.0%; Score 116; DB 9; Length 339;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REVHPLATSHOOYFYKIPILV 22
Db 92 REVHPLATSHOOYFYKIPILV 113

RESULT 10
US-10-011-095-2
; Sequence 2, Application US/10011095
; Publication No. US20030045682A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Hubert, Rene S.
; APPLICANT: Leong, Kahan
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Saifran, Douglas C.
; APPLICANT: Mitchell, Steve Chappell
; TITLE OF INVENTION: ANTIBODIES IMMUNOSPECIFIC FOR STEAP1 (AS AMENDED)
; FILE REFERENCE: 511582001610
; CURRENT APPLICATION NUMBER: US/10/011,095
; CURRENT FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: 09/323,873
; PRIOR FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-01
; PRIOR APPLICATION NUMBER: 60/091,183
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DNA
US-10-011-095-2

Query Match 100.0%; Score 116; DB 9; Length 339;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REVHPLATSHOOYFYKIPILV 22
Db 92 REVHPLATSHOOYFYKIPILV 113

RESULT 11
US-10-010-667A-2
; Sequence 2, Application US/10010667A
; Publication No. US20030055217A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
```

APPLICANT: Hubert, Rene S.  
APPLICANT: Leong, Kahan  
APPLICANT: Raitano, Arthur B.  
APPLICANT: Saffran, Douglas C.  
APPLICANT: Mitchell, Steve Chappell  
TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS  
TITLE OF INVENTION: EXPRESSED IN HUMAN CANCERS AND USES THEREOF  
FILE REFERENCE: 511582001601  
CURRENT APPLICATION NUMBER: US/10/010,667A  
CURRENT FILING DATE: 2001-12-06  
PRIOR APPLICATION NUMBER: 09/323,873  
PRIOR FILING DATE: 1998-06-01  
PRIOR APPLICATION NUMBER: 60/087,520  
PRIOR FILING DATE: 1998-06-01  
PRIOR APPLICATION NUMBER: 60/091,183  
PRIOR FILING DATE: 1998-06-30  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 339  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-010-667A-2

Query Match 100.0%; Score 116; DB 9; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1.7e-10;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 REVHPLATSHOOYFYKIPILV 22  
Db 92 REVHPLATSHOOYFYKIPILV 113

RESULT 12  
US-09-759-143-879  
Sequence 879, Application US/09759143  
Patent No. US2002002248A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Dillon, David C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Harlocker, Susan L.  
APPLICANT: Jiang, Yuqul  
APPLICANT: Henderson, Robert A.  
APPLICANT: Kalos, Michael D.  
APPLICANT: Fanger, Gary R.  
APPLICANT: Retter, Marc W.  
APPLICANT: Stolk, John A.  
APPLICANT: Day, Craig H.  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Carter, Darrick  
APPLICANT: Li, Samuel  
APPLICANT: Wang, Aijun  
APPLICANT: Skelky, Yasir A.W.  
APPLICANT: Hepler, William  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
FILE REFERENCE: 210121.427C23  
CURRENT APPLICATION NUMBER: US/09/759,143  
CURRENT FILING DATE: 2001-01-12  
NUMBER OF SEQ ID NOS: 934  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 879  
LENGTH: 339  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-759-143-879

Query Match 100.0%; Score 116; DB 10; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1.7e-10;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 REVHPLATSHOOYFYKIPILV 22

Db 92 REVHPLATSHOOYFYKIPILV 113

RESULT 13  
US-09-780-669-879  
Sequence 879, Application US/09780669  
Patent No. US20020051977A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Dillon, David C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Harlocker, Susan L.  
APPLICANT: Jiang, Yuqul  
APPLICANT: Henderson, Robert A.  
APPLICANT: Kalos, Michael D.  
APPLICANT: Fanger, Gary R.  
APPLICANT: Retter, Marc W.  
APPLICANT: Stolk, John A.  
APPLICANT: Day, Craig H.  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Carter, Darrick  
APPLICANT: Li, Samuel  
APPLICANT: Wang, Aijun  
APPLICANT: Skelky, Yasir A.W.  
APPLICANT: Hepler, William  
APPLICANT: Hural, John  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Houghton, Raymond L.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
FILE REFERENCE: 210121.427C24  
CURRENT APPLICATION NUMBER: US/09/780,669  
CURRENT FILING DATE: 2001-02-09  
NUMBER OF SEQ ID NOS: 943  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 879  
LENGTH: 339  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-780-669-879

Query Match 100.0%; Score 116; DB 10; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1.7e-10;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 REVHPLATSHOOYFYKIPILV 22  
Db 92 REVHPLATSHOOYFYKIPILV 113

RESULT 14  
US-09-822-827-879  
Sequence 879, Application US/09822827  
Patent No. US20020081680A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
FILE REFERENCE: 210121.534C1  
CURRENT APPLICATION NUMBER: US/09/822,827  
CURRENT FILING DATE: 2001-03-28  
NUMBER OF SEQ ID NOS: 982  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 879  
LENGTH: 339  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-822-827-879

Query Match 100.0%; Score 116; DB 10; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1.7e-10;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REVHPLATSHQYFYKIPILV 22  
 |||||  
 Db 92 REVHPLATSHQYFYKIPILV 113

RESULT 15

US-09-802-520-1  
 ; Sequence 1, Application US/09802520  
 ; Publication No. US20020187472A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lal, Preeti  
 ; APPLICANT: Faris, Mary  
 ; APPLICANT: Chen, Huel-Mei  
 ; APPLICANT: Ison, Craig H.  
 ; TITLE OF INVENTION: STEAP-RELATED PROTEIN  
 ; FILE REFERENCE: PC-0037 US  
 ; CURRENT APPLICATION NUMBER: US/09/802,520  
 ; CURRENT FILING DATE: 2001-03-09  
 ; NUMBER OF SEQ ID NOS: 11  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 1  
 ; LENGTH: 490  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; OTHER INFORMATION: Incyte ID No. US20020187472A1 7492448CD1  
 US-09-802-520-1

Query Match 58.6%; Score 68; DB 9; Length 490;  
 Best Local Similarity 65.0%; Pred. No. 0.0083;  
 Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 REVHPLATSHQYFYKIPILV 20  
 |||||  
 Db 233 RDVHPIYARNQSDPYKIPILV 252

Search completed: March 26, 2003, 17:04:15  
 Job time: 15.8372 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: March 26, 2003, 16:44:40 ; Search time 14.0698 Seconds  
(without alignments)  
46.007 Million cell updates/sec

Title: US-10-010-667A-19

Perfect score: 116

Sequence: 1 REVHPLATSHQYFYKIPILV 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

262574

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued\_Patents\_AA.\*

2: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*

3: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*

4: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep.\*

5: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*

6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	116	100.0	22	4	US-09-323-873A-19
2	116	100.0	339	4	US-09-323-873A-2
3	46	39.7	195	4	US-09-134-001C-3548
4	44	37.9	933	2	US-08-682-847-2
5	42	36.2	284	4	US-09-357-251-24
6	42	36.2	319	4	US-09-134-001C-3888
7	42	36.2	1140	4	US-08-471-112A-4
8	42	36.2	1809	3	US-09-012-515A-12
9	42	36.2	1809	4	US-08-360-144A-12
10	42	36.2	1809	4	US-09-012-504A-12
11	42	36.2	2549	5	US-08-471-112A-3
12	42	36.2	2549	5	PCT-US85-06722-12
13	41.5	35.8	625	3	US-08-581-148C-18
14	41.5	35.8	625	4	US-08-759-436-3
15	41.5	35.8	625	4	US-08-759-436-5
16	41	35.3	190	1	US-08-106-981-2
17	41	35.3	392	2	US-08-282-197C-60
18	41	35.3	396	4	US-09-461-474-2
19	41	35.3	592	2	US-08-366-490-2
20	41	35.3	592	3	US-08-860-483A-2
21	41	35.3	879	1	US-08-220-151-3
22	41	35.3	879	1	US-08-220-151-3
23	41	35.3	879	1	US-08-413-118-2
24	41	35.3	879	1	US-08-413-118-3
25	41	35.3	879	1	US-08-413-118-106
26	41	35.3	879	3	US-08-473-446-2
27	41	35.3	879	3	US-08-473-446-3

28	41	35.3	879	3	US-08-473-446-106	Sequence 106, App
29	40	34.5	68	4	US-09-134-001C-2858	Sequence 2858, Ap
30	40	34.5	115	2	US-07-903-029-6	Sequence 6, Appl1
31	40	34.5	206	4	US-09-219-797-2	Sequence 2, Appl1
32	40	34.5	303	1	US-08-185-437-5	Sequence 5, Appl1
33	40	34.5	353	4	US-09-461-474-4	Sequence 4, Appl1
34	40	34.5	485	4	US-09-291-023A-2	Sequence 2, Appl1
35	40	34.5	485	4	US-09-291-023A-4	Sequence 4, Appl1
36	40	34.5	485	4	US-09-291-023A-14	Sequence 14, Appl
37	40	34.5	485	4	US-09-291-023A-15	Sequence 15, Appl
38	40	34.5	566	4	US-09-411-628-4	Sequence 4, Appl1
39	40	34.5	737	1	US-08-185-437-2	Sequence 2, Appl1
40	40	34.5	737	1	US-08-185-437-4	Sequence 4, Appl1
41	40	34.5	868	1	US-08-220-151-7	Sequence 7, Appl1
42	40	34.5	868	1	US-08-413-118-7	Sequence 7, Appl1
43	40	34.5	868	3	US-08-804-439A-21	Sequence 21, Appl
44	40	34.5	868	3	US-08-473-446-7	Sequence 7, Appl1
45	40	34.5	868	3	US-08-720-229-21	Sequence 21, Appl

#### ALIGNMENTS

```
RESULT 1
US-09-323-873A-19
; Sequence 19, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahan Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; FILE REFERENCE: 129.16US02
; CURRENT APPLICATION NUMBER: US/09/323,873A
; PRIOR FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-01
; PRIOR APPLICATION NUMBER: 60/091,183
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: STEAP-1 PEPTIDE
US-09-323-873A-19

Query Match      100.0%; Score 116; DB 4; Length 22;
Best Local Similarity 100.0%; Pred.No. 4.3e-12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REVHPLATSHQYFYKIPILV 22
DB 1 REVHPLATSHQYFYKIPILV 22

RESULT 2
US-09-323-873A-2
; Sequence 2, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahan Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Steve Chappell Mitchell
```

```
/ TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
/ FILE OF INVENTION: EXPRESSED IN HUMAN CANCERS AND USES THEREOF
/ FILE REFERENCE: 129,16US02
/ CURRENT APPLICATION NUMBER: US/09/323,873A
/ CURRENT FILING DATE: 1999-06-01,520
/ PRIOR APPLICATION NUMBER: 60/087,520
/ PRIOR FILING DATE: 1998-06-01
/ PRIOR APPLICATION NUMBER: 60/091,183
/ PRIOR FILING DATE: 1998-06-30
/ NUMBER OF SEQ ID NOS: 32
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 2
/ LENGTH: 339
/ TYPE: PRT
/ ORGANISM: Homo Sapiens
/ US-09-323-873A-2

Query Match          100.0%; Score 116; DB 4; Length 339;
Best Local Similarity 100.0%; Pred. No. 9,1e-11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REVHPLATSHQGYFYKIPILV 22
Db 92 REVHPLATSHQGYFYKIPILV 113

RESULT 3
US-09-134-001C-3548
/ Sequence 3548, Application US/09134001C
/ Patent No. 6380370
/ GENERAL INFORMATION:
/ APPLICANT: Lynn Doucette-Stamm et al
/ TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
/ TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: GTC-007
/ CURRENT APPLICATION NUMBER: US/09/134,001C
/ CURRENT FILING DATE: 1998-08-13
/ PRIOR APPLICATION NUMBER: US 60/064,964
/ PRIOR FILING DATE: 1997-11-08
/ PRIOR APPLICATION NUMBER: US 60/055,779
/ PRIOR FILING DATE: 1997-08-14
/ NUMBER OF SEQ ID NOS: 5674
/ SEQ ID NO 3548
/ LENGTH: 195
/ TYPE: PRT
/ ORGANISM: Staphylococcus epidermidis
/ US-09-134-001C-3548

Query Match          39.7%; Score 46; DB 4; Length 195;
Best Local Similarity 44.4%; Pred. No. 6;
Matches 8; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 REVHPLATSHQGYFYKIP 18
Db 128 RSLSPKTSYQEWYOI 145

RESULT 4
US-08-682-847-2
/ Sequence 2, Application US/08682847
/ Patent No. 5856989
/ GENERAL INFORMATION:
/ APPLICANT: BABIUK, LORNE
/ APPLICANT: VAN DEN HURK, SYLVIA
/ APPLICANT: ZAMB, TIM
/ APPLICANT: FITZPATRICK, DAVID
/ TITLE OF INVENTION: RECOMBINANT BOVINE HERPESVIRUS TYPE 1
/ TITLE OF INVENTION: POLYPEPTIDES AND VACCINES
/ NUMBER OF SEQUENCES: 8
/ CORRESPONDENCE ADDRESSES:
/ ADDRESSEE: MORRISON & FOERSTER
/ STREET: 755 PAGE MILL ROAD
/ CITY: PALO ALTO
```

```
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94304-1018
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/682,847
/ FILING DATE: 12-JUL-1996
/ CLASSIFICATION: 536
/ ATTORNEY/AGENT INFORMATION:
/ NAME: PARK, FREDIE K.
/ REGISTRATION NUMBER: 35,636
/ REFERENCE/DOCKET NUMBER: 29310-20005.10
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 813-5600
/ TELEFAX: (415) 494-0792
/ TELEX: 706141
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 933 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-682-847-2

Query Match          37.9%; Score 44; DB 2; Length 933;
Best Local Similarity 40.0%; Pred. No. 71;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 REVHPLATSHQGYFYKIP 15
Db 660 RELVEPCTANHKRYF 674

RESULT 5
US-09-357-251-24
/ Sequence 24, Application US/09357251
/ Patent No. 6271441
/ GENERAL INFORMATION:
/ APPLICANT: Falco, S. Carl
/ APPLICANT: Famodu, Layo O.
/ APPLICANT: Orozco, Buddy
/ APPLICANT: Schwaber, James S.
/ TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase
/ FILE REFERENCE: BB-1193
/ CURRENT APPLICATION NUMBER: US/09/357,251
/ CURRENT FILING DATE: 1999-07-20
/ EARLIER APPLICATION NUMBER: 60/093,530
/ EARLIER FILING DATE: July 21, 1998
/ NUMBER OF SEQ ID NOS: 37
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 24
/ LENGTH: 284
/ TYPE: PRT
/ ORGANISM: Trifolium aestivum
/ US-09-357-251-24

Query Match          36.2%; Score 42; DB 4; Length 284;
Best Local Similarity 46.7%; Pred. No. 39;
Matches 7; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 5 HPLATSHQGYFYKIP 19
Db 55 HPARDSDHTFLKAP 69

RESULT 6
US-09-134-001C-3888
/ Sequence 3888, Application US/09134001C
/ Patent No. 6380370
```



```

; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3888
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3888

Query Match          36.2%; Score 42; DB 4; Length 319;
Best Local Similarity 47.1%; Pred. No. 45;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Qy 4 IHPLATSHOOQFYKIP1 20
Db 249 IHKXCSDHOQRFYNC51 265

RESULT 7
US-08-471-112A-4
; Sequence 4, Application US/08471112A
; Patent No. 6313264
; GENERAL INFORMATION:
; APPLICANT: Molnar-Kimber, Katherine L.
; APPLICANT: Caggiano, Thomas F.
; APPLICANT: Nakanishi, Koji
; APPLICANT: Chen, Yanqiu
; TITLE OF INVENTION: EFFECTOR PROTEINS OF RAPAMYCIN
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,112A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/384,524
; FILING DATE: 13-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/312,023
; FILING DATE: 26-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/207,975
; FILING DATE: 08-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Siekman, Michael T.
; REGISTRATION NUMBER: 36,276
; REFERENCE/DOCKET NUMBER: 01142.0058-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 4:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-471-112A-4

Query Match          36.2%; Score 42; DB 4; Length 1140;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 REVIHPLATSHOOQFY 16
Db 274 RQLDHPLPTVHPQVY 289

RESULT 8
US-09-012-515A-12
; Sequence 12, Application US/09012515A
; Patent No. 6127521
; GENERAL INFORMATION:
; APPLICANT: Berlin, Vivian
; APPLICANT: Chiu, Maria Isabel
; APPLICANT: Cottarel, Guillaume
; APPLICANT: Damaguez, Veronique
; TITLE OF INVENTION: IMMUNOSUPPRESSANT TARGET PROTEINS
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HONG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/012,515A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/360,144
; FILING DATE: 20-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: APV-036.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1809 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-012-515A-12

Query Match          36.2%; Score 42; DB 3; Length 1809;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 REVIHPLATSHOOQFY 16
Db 943 RQLDHPLPTVHPQVY 958

RESULT 9
US-08-360-144A-12

```

Sequence 12, Application US/08360144A  
Patent No. 6150137  
GENERAL INFORMATION:  
APPLICANT: Berlin, Vivian  
APPLICANT: Chiu, Maria Isabel  
APPLICANT: Cottarel, Guillaume  
APPLICANT: Damagnez, Veronique  
TITLE OF INVENTION: IMMUNOSUPPRESSANT TARGET PROTEINS  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/360,144A  
FILING DATE: 20-DEC-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Vincent, Matthew P.  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: APV-036.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1809 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-360-144A-12

Query Match 36.2%; Score 42; DB 4; Length 1809;  
Best Local Similarity 50.0%; Pred. No. 3.1e+02;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 REVHPLATSHOOQFY 16  
DB 943 ROLDHPLPTVHPQVTV 958

RESULT 10  
US-09-012-504A-12  
Sequence 12, Application US/09012504A  
Patent No. 6464974  
GENERAL INFORMATION:  
APPLICANT: Berlin, V.  
APPLICANT: Chiu, I.  
APPLICANT: Cottarel, G.  
APPLICANT: Damagnez, V.  
TITLE OF INVENTION: IMMUNOSUPPRESSANT TARGET PROTEINS  
FILE REFERENCE: APEI-P05-036  
CURRENT APPLICATION NUMBER: US/09/012,504A  
CURRENT FILING DATE: 1998-01-23  
PRIOR APPLICATION NUMBER: 08/360,144  
PRIOR FILING DATE: 1994-12-20  
PRIOR APPLICATION NUMBER: 08/250,795  
PRIOR FILING DATE: 1994-05-27  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 12  
LENGTH: 1809  
TYPE: PRT  
ORGANISM: Mammalian  
US-09-012-504A-12

Query Match 36.2%; Score 42; DB 4; Length 1809;  
Best Local Similarity 50.0%; Pred. No. 3.1e+02;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 REVHPLATSHOOQFY 16  
DB 943 ROLDHPLPTVHPQVTV 958

RESULT 11  
US-08-471-112A-3  
Sequence 3, Application US/08471112A  
Patent No. 6313264  
GENERAL INFORMATION:  
APPLICANT: Molnar-Kimber, Katherine L.  
APPLICANT: Fallis, Amedeo F.  
APPLICANT: Cagliano, Thomas J.  
APPLICANT: Nakanishi, Koji  
APPLICANT: Chen, Yanqiu  
TITLE OF INVENTION: EFFECTOR PROTEINS OF RAPAMYCIN  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner, L.L.P.  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,112A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/384,524  
FILING DATE: 13-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/312,023  
FILING DATE: 26-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/207,975  
FILING DATE: 08-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Sleiman, Michael T.  
REGISTRATION NUMBER: 36,276  
REFERENCE/DOCKET NUMBER: 01142.0058-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2549 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-471-112A-3

Query Match 36.2%; Score 42; DB 4; Length 2549;  
Best Local Similarity 50.0%; Pred. No. 4.5e+02;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 REVHPLATSHOOQFY 16  
DB 1683 ROLDHPLPTVHPQVTV 1698

```
RESULT 12
PCT-US95-06722-12
; Sequence 12, Application PC/TUS9506722
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Immunosuppressant Target Proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06722
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/250,795
; FILING DATE: 27-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/250,795
; FILING DATE: 20-DEC-1994
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2549 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-06722-12

Query Match          36.2%; Score 42; DB 5; Length 2549;
Best Local Similarity 50.0%; Pred. No. 4.5e+02;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy      1 REVHPLATSHQOYF 16
Db      1683 RQLDHPLPTVHPQVTV 1698

RESULT 13
US-08-581-148C-18
; Sequence 18, Application US/08581148C
; Patent No. 6060644
; GENERAL INFORMATION:
; APPLICANT: Schnable, Patrick S.
; APPLICANT: Robertson, Donald S.
; APPLICANT: Hansen, Joel D.
; APPLICANT: Nikolau, Basil J.
; APPLICANT: Xu, Xiaojie
; APPLICANT: Xia, Yiji
; TITLE OF INVENTION: ISOLATION AND USE OF CUTICULAR LIPID
; TITLE OF INVENTION: GENES
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Leydig, Volt & Mayer, Ltd.
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL
; COUNTRY: US
; ZIP: 60601-6780
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/581,148C
; FILING DATE: 29-DEC-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Larcher, Carol
; REGISTRATION NUMBER: 35243
; REFERENCE/DOCKET NUMBER: 71380
```

```
TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5600
; TELEFAX: (312) 616-5700
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 625 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-581-148C-18

Query Match          35.8%; Score 41.5; DB 3; Length 625;
Best Local Similarity 52.4%; Pred. No. 1.1e+02;
Matches 11; Conservative 2; Mismatches 5; Indels 3; Gaps 2;

Qy      3 VIHPLATSHQOYF--YKIPIL 21
Db      174 VIHPPA-EHIAVFIPLFAIPPL 193

RESULT 14
US-08-759-436-3
; Sequence 3, Application US/08759436
; Patent No. 6437218
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DNA fragment encoding a protein
; TITLE OF INVENTION: involved in
; TITLE OF INVENTION: molecules comprising said fragment and a method for
; TITLE OF INVENTION: obtaining transformed bacterial cells and plants
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Iadec & Parry
; STREET: 26 West 61st Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10023
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/759,436
; FILING DATE: 5-Dec-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Cord, Janet I.
; REGISTRATION NUMBER: 33,778
; REFERENCE/DOCKET NUMBER: U011063-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-708-1935
; TELEFAX: 212-246-8959
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 625 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-759-436-3

Query Match          35.8%; Score 41.5; DB 4; Length 625;
Best Local Similarity 52.4%; Pred. No. 1.1e+02;
Matches 11; Conservative 2; Mismatches 5; Indels 3; Gaps 2;

Qy      3 VIHPLATSHQOYF--YKIPIL 21
Db      174 VIHPPA-EHIAVFIPLFAIPPL 193

RESULT 15
US-08-759-436-5
```

Search completed: March 26, 2003, 16:51:39  
Job time : 16.0698 secs

Sequence 5, Application US/08759436  
Patent No. 6437218  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: DNA fragment encoding a protein  
TITLE OF INVENTION: Involved in  
TITLE OF INVENTION: fatty aldehyde decarboxylase activity, recombinant  
TITLE OF INVENTION: molecules comprising said fragment and a method for  
TITLE OF INVENTION: obtaining transformed bacterial cells and plants  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ladas & Parry  
STREET: 26 West 61st Street  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10023  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/759,436  
FILING DATE: 5-Dec-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Cord, Janet I.  
REGISTRATION NUMBER: 33,778  
REFERENCE/DOCKET NUMBER: U011063-5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-708-1935  
TELEFAX: 212-246-8959  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 625 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Arabidopsis thaliana  
STRAIN: Columbia ecotype  
DEVELOPMENTAL STAGE: light and dark grown seedlings;  
DEVELOPMENTAL STAGE: rosette and whole plants  
IMMEDIATE SOURCE:  
LIBRARY: PRL2  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..625  
PUBLICATION INFORMATION:  
AUTHORS: Aarts, Mark G.M.  
AUTHORS: Keljzer, Christiaan J.  
AUTHORS: Stiekema, Willem J.  
AUTHORS: Pereira, Andy  
TITLE: Molecular characterization of the CER1 gene  
TITLE: of Arabidopsis involved in epicuticular wax  
TITLE: biosynthesis and pollen fertility  
JOURNAL: Plant Cell  
VOLUME: 7  
ISSUE: 12  
PAGES: 2115-2127  
DATE: december-1995  
RELEVANT RESIDUES IN SEQ ID NO: 5: FROM 1 TO 625  
US-08-759-436-5

Query Match 35.8%; Score 41.5; DB 4; Length 625;  
Best Local Similarity 52.4%; Pred. No. 1,1e+02;  
Matches 11; Conservative 2; Mismatches 5; Indels 3; Gaps 2;

QY 3 VHPPLATSHOOF--YKIPIL 21  
DB 174 VHPFA-EHIVFPLFAIPPL 193